

CVPR
JUNE 3-7, 2026



DENVER
COLORADO

**IEEE/CVF Conference on
Computer Vision and Pattern Recognition**



WORKSHOPS & TUTORIALS PROGRAM

ULTIMATE SPONSORS

build



Mohamed bin Zayed
University of
Artificial Intelligence

Institute of
Foundation Models

Lambda

Meta

Tencent

T E S L F

Uber

WAYMO

PLATINUM SPONSORS

Adobe

Alibaba

amazon

ANT GROUP



ByteDance

MINIMAX

ORACLE

Qualcomm

SONY

TORC

GOLD SPONSORS

Baidu

baseten

contra LABS

Crusoe

DYNA

ENCORD

Luma

美团 美团

NVIDIA

TURING

Matterport

UT PAI

Weights & Biases
by CoreWeave

SILVER SPONSORS

Appen

Applied Intuition

Aurora

BagelLABS



GMI

HPC-AI.COM

Ideogram

kitware

LINKERB

Motional

nexdata #sudo

VOXEL51

waabi

ZOOX

Welcome to the 2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition in Denver, Colorado, the “Mile High City,” located at an elevation of 5280 feet (1 mile)! CVPR is the premier and flagship annual meeting of IEEE/CVF and PAMI-TC, where researchers in our community present their latest advances in computer vision, pattern recognition, machine learning, robotics, and artificial intelligence, both in theory and practice. CVPR 2026 is primarily an in-person conference, but for those unable to attend in person, we are pleased to offer a virtual component with access to conference papers, posters, videos, and talks.

CVPR 2026 received 16,092 valid paper submissions that underwent the review process, a 24% increase from CVPR 2025. The review process was managed by the 6 Program Co-Chairs, 41 Senior Area Chairs, and 909 Area Chairs. During the review phase, each paper received at least 3 reviews from a pool of 25,149 reviewers. As in prior years, after receiving these initial reviews, authors had the opportunity to submit a rebuttal to the reviews. The process concluded with discussions among reviewers and ACs, finalizing of reviews, and ACs working in triplets to make final accept/reject decisions for each paper. At the end of this process, 4,089 papers were accepted, for a 25.4% overall acceptance rate. In keeping with the CVPR tradition, the PCs did not pre-define any target acceptance rate or number of papers to be accepted; the resulting acceptance rate reflects the community consensus and is consistent with past CVPRs.

All the 4,089 accepted papers were invited to present posters at CVPR. In addition, 141 (3.4%) papers were selected to be presented as oral talks, based on nominations from Area Chairs, and 578 (14.1%) papers were selected by ACs together with Senior ACs to be “highlights” due to their high quality and potential impact. This year, the oral presentations will be organized into four tracks. The highlights are also flagged with a special annotation in the program. 74 papers were nominated to be the best paper award candidates, from which a committee convened by the PCs selected the award winners to be announced during the conference.

Due to the dramatic increase in submissions and a commitment to improving the quality and accountability of the review process, the PCs adopted several measures introduced in recent CVPR review processes. These measures include imposing a limit on the number of submissions per author, requiring authors to participate in the review process when deemed qualified by the PCs, and rejecting papers submitted by highly irresponsible reviewers. In addition, we also adopted the outstanding reviewer recognition, highlighting 1,545 reviewers (6.1%) from our large pool of 25,149 reviewers. Papers authored by outstanding reviewers are flagged with a special annotation in the program. We further introduced the outstanding AC recognition, highlighting 303 ACs (33.3%) from our pool of 909 ACs for their timely and outstanding service.

New this year, CVPR 2026 introduces an experimental Compute Reporting Initiative, inviting authors to disclose the computational resources used in their work. This initiative reflects our community’s growing interest in understanding the computational landscape of CV/AI research, from resource requirements to efficiency innovations. Compute reports were not visible to reviewers and did not influence acceptance decisions. We hope this data will enable community-wide benchmarking and inspire future work on computational efficiency and accessibility. Awards will be presented recognizing exceptional efficiency and transparency. Highlights from this initiative will be presented during the conference, and a full report will be disseminated to the community thereafter.

We would like to thank everyone involved in making CVPR 2026 a success. This includes the organizing committee, Senior Area Chairs, Area Chairs, reviewers, authors, demo session participants, contributing artists, and donors and exhibitors. It is exciting to be part of a team of people around the world united in a common goal. The scale and quality of this year’s program is a testament to the dedication of this global community.

We also thank Nicole Finn and her C to C Events team for organizing the conference logistics, Yoshitomo Matsubara for his amazing work as technical chair, Lee Campbell and the Event Hosts team for their work on the website and virtual platform, Mike Weil and Hall Erickson for handling sponsorships and the exhibition, and Luba Elliot as our AI Art Coordinator. Finally, we thank all of you for attending CVPR 2026 and making it one of the top venues for computer vision research in the world. We hope that you also have some time to explore Denver during the conference.

Enjoy CVPR 2026. We look forward to meeting you in person!

CVPR Program Chairs

Angela Dai (*TU Munich*)

Adriana Kovashka (*University of Pittsburgh*)

Chen Change Loy (*Nanyang Technological University*),

Vladimir Pavlovic (*Rutgers University*),

Alex Schwing (*University of Illinois Urbana-Champaign*),

Shaoting Zhang (*Shanghai Jiao Tong University*)

CVPR 2026 General Chairs

Chandra Kambhampettu (*University of South Florida*)

Dimitris N. Metaxas (*Rutgers University*)

CVPR 2026 Virtual Platform



Access schedules,
papers, workshops,
tutorials, etc

Exhibit Hall Floorplan



Access location of
exhibitors on the
main floor

Honorary General Chairs

Takeo Kanade (*CMU*)
Shimon Ullman (*Weizmann Institute of Science*)

General Chairs

Chandra Kambhampettu (*University of South Florida*)
Dimitris N. Metaxas (*Rutgers University*)

Program Chairs

Angela Dai (*Technical University of Munich*)
Adriana Kovashka (*University of Pittsburgh*)
Chen Change Loy (*Nanyang Technological University*)
Vladimir Pavlovic (*Rutgers University*)
Alexander G. Schwing (*University of Illinois Urbana-Champaign*)
Shaoting Zhang (*Shanghai Jiao Tong University*)

Advisor to the Program Committee

David Forsyth (*University of Illinois Urbana-Champaign*)

Workshop Chairs

Erik Blasch (*AFRL*)
Mei Chen (*Dolby*)
Brian Clipp (*Motional*)
Sharon X. Huang (*Penn State University*)
Andreas Savakis (*Rochester Institute of Technology*)
Humphrey Shi (*Georgia Tech*)

Tutorial Chairs

Boqing Gong (*Boston University / Google*)
Hongsheng Li (*The Chinese University of Hong Kong*)
Paul Schrader (*Air Force Research Laboratory*)
Vitomir Struc (*University of Ljubljana*)

Demonstration Chairs

Sathyanarayanan N. Aakur (*Auburn University*)
Shu Kong (*University of Macau*)

Finance Chair

Walter Scheirer (*Notre Dame*)

Publicity Chairs

Deblina Bhattacharjee (*University of Bath*)
Kosta Derpanis (*York University*)
Antonino Furnari (*University of Catania*)
Zhengzhong Tu (*Texas A&M University, UT Austin*)
Yael Vinker (*MIT*)

Broadening Participation Chairs

Naresh Cuntoor (*BlueHalo Labs*)
Christopher Funk (*Kitware*)
Deepti Ghadiyaram (*Boston University*)
Michael King (*Florida Tech*)
Roni Sengupta (*UNC Chapel Hill*)

Senior PAMI-TC Ombuds

David Forsyth (*University of Illinois Urbana-Champaign*)
Linda Shapiro (*University of Washington*)

AI Art Curator

Luba Elliott (*Independent Curator*)

Technical Chair

Yoshitomo Matsubara (*Yahoo!*)

Doctoral Consortium Chairs

Paola Cascante-Bonilla (*Stonybrook University*)
Abby Stylianou (*St. Louis University*)

Local Chair

Brendan Klare (*ROC*)

Web Developer

Lee Campbell (*Eventhosts*)

Conference Producer

Nicole Finn (*c to c events*)

Compute Reporting Committee Chairs

Sourav Bhattacharya (*NatWest Group*)
Christopher Funk (*Kitware, Inc.*)
Farhad Pourkamali (*University of Colorado Denver*)
Roy Schwartz (*Hebrew University of Jerusalem*)
Emma Strubell (*Carnegie Mellon University*)
Sasha Luccioni (*Hugging Face*) - Advisor

Award Committee

Derek Hoiem (*Chair, University of Illinois Urbana-Champaign*)
Kristen Grauman (*UT Austin*)
Ko Nishino (*Kyoto University*)
Deqing Sun (*Google Deepmind*)
Tinne Tuytelaars (*KU Leuven*)
Angela Yao (*National University of Singapore*)
Jun-Yan Zhu (*Carnegie Mellon University*)

Oral Coaching Committee

Srinivasa Narasimhan (*Chair, Carnegie Mellon University*)
Dima Damen (*University of Bristol*)
Fernando De la Torre (*Carnegie Mellon University*)
Sven Dickinson (*University of Toronto*)
Alexei Efros (*UC Berkeley*)
David Forsyth (*University of Illinois Urbana-Champaign*)
Bill Freeman (*MIT*)
Ioannis Gkioulekas (*Carnegie Mellon University*)
Mohit Gupta (*University of Wisconsin, Madison*)
Kiriakos Kutulakos (*University of Toronto*)
Jean-Francois Lalonde (*Université Laval*)
Lana Lazebnik (*University of Illinois Urbana-Champaign*)
Ko Nishino (*Kyoto University*)
Devi Parikh (*Yutori*)
Deva Ramanan (*Carnegie Mellon University*)
Aswin Sankaranarayanan (*Carnegie Mellon University*)
Steve Seitz (*University of Washington, Google*)
Yaser Sheikh (*Sooth Labs*)
Ashok Veeraraghavan (*Rice University*)
Yair Weiss (*Hebrew University of Jerusalem*)
Richard Zhang (*Adobe Research*)
Todd Zickler (*Harvard University*)

To make it easier to navigate the growing number of CVPR workshops, we have grouped the workshops together into thematic “tracks.” The workshops within each track cover closely related topics, and we’ve also tried our best to avoid having scheduling conflicts between them. We hope you enjoy the workshops!

Workshop Chairs

Erik Blasch, Mei Chen, Brian Clipp, Sharon X. Huang, Andreas Savakis, Humphrey Shi

TABLE OF CONTENTS

Tracks and Workshops	Date	Time	Location	Page
Track on 3D from Multi-view and Sensors				
2nd Workshop on 4D Vision: Modeling the Dynamic World	6/4/2026...	PM.....	506	40
3rd Workshop on ScanNet++ Novel View Synthesis and 3D Semantic Understanding Challenge	6/3/2026...	PM.....	710	26
Eighth Workshop on Image Matching: Local Features and Beyond.....	6/4/2026...	PM.....	504	43
SPAR-3D: Security, Privacy, and Adversarial Robustness in 3D Generative Vision Models	6/4/2026...	AM.....	Mile High 3A	34
Spatial Intelligence for Cultural Heritage.....	6/3/2026...	PM.....	708	26
Third Workshop for Learning 3D with Multi-View Supervision	6/4/2026...	AM.....	703	30
Urban Scene Modeling: Structured, Semantic, and Synthetic 3D Habitats.....	6/3/2026...	Full Day.....	Mile High 3B	20
Track on Adversarial Attack and Defense				
Synthetic & Adversarial ForEnsis.....	6/3/2026...	PM.....	107	26
The 6th Workshop of Adversarial Machine Learning on Computer Vision: Safety of Vision-Language Agents	6/4/2026...	Full Day.....	708	35
Track on Affinity Groups				
LatinX in Computer Vision Research Workshop	6/3/2026...	AM.....	106	14
Women in Computer Vision	6/3/2026...	AM.....	708	16
Track on Autonomous Driving				
9th International Workshop on Visual Odometry and Computer Vision Applications Based on Location Clues	6/4/2026...	AM.....	506	35
Autonomous Understanding Through Open-world Perception and Integrated Language models for On-road Tasks.....	6/3/2026...	Full Day.....	Mile High 3A	18
Foundation Models for V2X-Based Cooperative Autonomous Driving.....	6/3/2026...	Full Day.....	207	18
Multi-Agent Embodied Intelligent Systems Meet Agentic-AI era: Opportunities, Challenges and Futures	6/4/2026...	Full Day.....	Four Seasons 1	38
The 1st Workshop on Deployment of Foundation Models for Embodied AI.....	6/3/2026...	Full Day.....	Four Seasons 2	20
The Eighth Workshop on Precognition: Seeing through the Future	6/4/2026...	AM.....	210/212	34
Third Workshop on Simulation for Autonomous Driving	6/4/2026...	PM.....	102/104	44
Workshop on Autonomous Driving.....	6/3/2026...	Full Day.....	603	20
Track on Biometrics				
AERO-HPR: Human Perception and Recognition in Aerial Surveillance	6/3/2026...	AM.....	110	12
CVPR 2026 Biometrics Workshop	6/4/2026...	PM.....	106	
Second Workshop on Foundation and Generative Models in Biometrics.....	6/3/2026...	PM.....	113	24
The 2nd International Workshop & Challenge on Subtle Visual Computing @CVPR 2026	6/3/2026...	AM.....	109	16

Track on Computational Imaging

11th New Trends in Image Restoration and Enhancement Workshop and Challenges	6/4/2026 ... Full Day	207	38
Computational Cameras and Displays	6/3/2026 ... Full Day	Mile High 4CD	18
The 8th UG2+ Workshop and Challenge: Bridging the Gap between Computational Photography and Visual Perception	6/4/2026 ... AM	Mile High 4EF	34

Track on Data/Synthetic Data for Vision

Auto-Annotation with Expert-Crafted Guidelines	6/3/2026 ... PM	711	22
Exploring the Next Generation of Data	6/4/2026 ... AM	603	32
The 3rd Workshop on Synthetic Data for Computer Vision	6/4/2026 ... PM	607	45
The 5th DataCV Workshop and Challenge	6/3/2026 ... AM	710	13

Track on Deep Learning Architecture and Techniques

13th Workshop on Fine-grained Visual Categorization	6/3/2026 ... Full Day	504	19
Domain Generalization: Evolution, Breakthroughs, and Future Horizons (2nd Edition)	6/4/2026 ... PM	103	42
Sixth Workshop on Neural Architecture Search	6/3/2026 ... PM	110	23
The 5th Workshop on Federated Learning for Computer Vision	6/3/2026 ... AM	711	13
The 5th Workshop on Transformers for Vision and Multimodal AI	6/3/2026 ... PM	607	26

Track on Efficient, Edge, and Scalable Vision

3rd Workshop on Efficient and On-Device Generation (EDGE), CVPR 2026	6/3/2026 ... PM	203	23
Efficient Deep Learning for Computer Vision	6/3/2026 ... Full Day	502	19
Mobile AI workshop and associated challenges, 6th edition	6/4/2026 ... Full Day	505	37
On Sensor Vision Workshop	6/3/2026 ... AM	Mile High 4EF	15
The 22nd Embedded Vision Workshop	6/3/2026 ... AM	709	13

Track on Embodied Vision: Active Agents, Simulation

1st Workshop on Multi-Agent Robotic Systems: Scaling with Compositional Intelligence	6/3/2026 ... PM	Mile High 4EF	24
2nd Workshop on Agents in Interaction, from Humans to Robots	6/4/2026 ... Full Day	Mile High 2A	37
4D Digital Twins: Real-to-Sim-to-Real for Physical AI	6/4/2026 ... PM	Mile High 2C	40
Bridging Vision, Language, and Action: What's Missing in Actionable Visual Perception for Robotics	6/3/2026 ... Full Day	503	18
Embodied Reasoning in Action: Workshop and Challenge on Embodied Reasoning for Robotic Manipulation	6/4/2026 ... Full Day	605	36
From Lab Demos to Daily Tasks: Embodied Intelligence in the Wild	6/3/2026 ... Full Day	Four Seasons 1	19
IPA: Interactive Physical AI Workshop	6/3/2026 ... AM	203	14
Sense of Space: Multi-Sensory Modeling for Embodied Intelligence	6/3/2026 ... Full Day	Mile High 2C	20

Track on Emerging Topics - Other

CV4Science: Using Computer Vision for the Sciences	6/4/2026 ... PM	709	42
Multimodal Alignment for a Pluralistic Society	6/3/2026 ... AM	113	15
Rediscovering Intelligence: Can AI Still Learn from Humans?	6/3/2026 ... PM	Mile High 4AB	25
Workshop on "Bitter Lessons"	6/3/2026 ... AM	Four Seasons 4	13

Track on Explainable Computer Vision

How Do Vision Models Work?	6/4/2026 ... Full Day	Mile High 1EF	37
Safe Artificial Intelligence for All Domains	6/4/2026 ... AM	103	34
The 5th Explainable AI for Computer Vision (XAI4CV) Workshop	6/3/2026 ... AM	Mile High 1AB	17

Track on Foundation Models (LLM, VLM, VLA, etc.)

2nd Workshop on Video Large Language Models 6/4/2026... Full Day..... Four Seasons 4 39

Big Model Adaptation In Computer Vision 6/4/2026... PM..... Four Seasons 2 41

DataMFM: Emerging Directions in Data for Multimodal Foundation Models 6/3/2026... PM..... 111 23

GigaBrain Challenge 2026: Workshop on
World Models Empowering Vision Language Action Model..... 6/3/2026... PM..... 507 14

ScaleBot: The First Workshop on Scalable Robot Learning Systems 6/4/2026... PM..... 610/612 45

The 2nd 3D-LLM/VLA Workshop:
Bridging Language, Vision and Action in 3D Environments 6/3/2026... PM..... Mile High 1CD 21

The 2nd CVPR Workshop Proposal
on Foundation Models Meet Embodied Agents 6/4/2026... PM..... 703 42

The 2nd Workshop on Multimodal Spatial Intelligence 6/3/2026... AM 601 15

The 5th Workshop on Computer Vision in the Wild:
Towards Unified Multimodal Agents For Reasoning in the Wild 6/4/2026... AM..... Four Seasons 2 31

The Second Workshop on the Evaluation
of the Generative Foundation Models 6/4/2026... AM..... Mile High 2C 31

Visual General Intelligence..... 6/3/2026... Full Day..... 703 20

Workshop on Agentic AI for Visual Media 6/3/2026... Full Day..... Mile High 1EF 18

Track on Generative Models

1st Workshop on Generative 3D Reconstruction 6/4/2026... PM..... 603 39

2nd Workshop on GenAI for Storytelling 6/4/2026... PM..... 105 41

3D Geometry Generation for Scientific Computing (2nd Edition) 6/3/2026... AM..... 105 30

4th Workshop on Generative Models for Computer Vision..... 6/4/2026... Full Day..... 205 37

Generative AI for XR and Identity-based Applications..... 6/3/2026... AM..... 111 14

Personalization in Generative AI Workshop 6/4/2026... AM..... Mile High 4CD 33

The 1st Workshop on Low-Level Vision Frontiers
with Generative AI, Preference Optimization, and Agentic Systems 6/4/2026... AM..... 504 32

Video Generative Models: Benchmarks and Evaluation 6/4/2026... Full Day..... Mile High 3B 38

Track on Human Modeling & Understanding: Face, Body, Pose, Gesture, Movement

10th Affective & Behavior Analysis in-the-wild 6/3/2026... PM..... 505 21

2nd Workshop on Computer Vision for Children 6/4/2026... AM..... 106 31

2nd Workshop on Human-Interactive Generation and Editing..... 6/4/2026... Full Day..... 109 37

2nd Workshop on Photorealistic 3D Head Avatars 6/3/2026... AM..... 107 16

Computer Vision for Biomechanics Workshop..... 6/3/2026... PM..... 112 23

Humans of Generative AI 6/4/2026... AM..... 710 32

PhysHuman: Physically Grounded Human Perception and Modeling..... 6/4/2026... AM..... 110 33

The 3rd Workshop on Human Motion Generation -
New Perspective on Simulation, Animation, and VR applications 6/3/2026... AM..... 505 14

The 7th International Workshop on Eye and Gaze in Computer Vision 6/4/2026... PM..... 711 43

Workshop on Multimodal Human Motion Analysis..... 6/3/2026... PM..... 601 24

Track on Image and Video Synthesis and Generation

1st Workshop on Journey to the Awards:
Generative AI for Movie-Grade Video Production (J2A)..... 6/4/2026... PM..... Mile High 4CD 43

1st Workshop on Video World Models: Interaction, Memory, and Efficiency 6/3/2026... AM..... 705/707 16

AI for Creative Visual Content Generation, Editing and Understanding 6/3/2026... AM..... 501 13

The 1st Workshop on AI-assisted Long Video Creation 6/3/2026... PM..... 712 21

Track on Medical and Biological Vision, Cell Microscopy

11th Workshop on Computer Vision and Multimodal Microscopy Image Analysis ...	6/4/2026 ... Full Day	112	36
Bridging AI and Medical Reality: Computer Vision for Real-world Clinical Translation	6/4/2026 ... PM	Mile High 1AB	42
Medical Reasoning with Vision Language Foundation Models	6/4/2026 ... PM	Mile High 1AB	40
Multimodal Foundation Models for Biomedicine: Challenges and Opportunities	6/3/2026 ... AM	Mile High 1CD	15
PHAROS AI Factory for Medical Imaging & Healthcare	6/3/2026 ... AM	108	17
12th Workshop on Medical Computer Vision	6/3/2026 ... Full Day	605	19
The 3rd Workshop on Foundation Models for Medical Vision	6/4/2026 ... PM	607	17

Track on Multimodal Detection, Recognition, Segmentation

22nd Workshop on Perception Beyond the Visible Spectrum	6/3/2026 ... AM	712	16
4th Workshop on Vision Based Industrial Inspection	6/3/2026 ... Full Day	205	20

Track on Multimodal Learning

9th Multimodal Learning and Applications Workshop	6/4/2026 ... Full Day	111	38
Sight and Sound	6/4/2026 ... Full Day	Mile High 1CD	39
The 5th Workshop on "What is Next in Multimodal Foundation Models?"	6/3/2026 ... PM	Four Seasons 4	24
Workshop on Any-to-any Multimodal Learning	6/4/2026 ... AM	502	30

Track on Open World Learning

Open-World Vision	6/4/2026 ... AM	712	33
The Third Workshop on Anomaly Detection with Foundation Models	6/4/2026 ... PM	712	40
Visual Anomaly and Novelty Detection - 4th Edition	6/4/2026 ... PM	601	45

Track on Photogrammetry and Remote Sensing

EarthVision: Large Scale Computer Vision for Remote Sensing Imagery	6/4/2026 ... Full Day	507	36
The 1st Workshop on Monitoring the World through an Imperfect Lens	6/3/2026 ... PM	105	24
The Second CVPR Workshop on Foundation and Large Vision Models in Remote Sensing (MORSE)	6/3/2026 ... PM	106	15

Track on Robot Perception

4th Workshop on Maritime Computer Vision	6/4/2026 ... AM	Mile High 1AB	35
Unified Robotic Vision with Cross-Modal Sensing and Alignment	6/4/2026 ... AM	Mile High 4AB	35
Workshop on World Models Meet Active Sensing and Closed-Loop Planning	6/3/2026 ... AM	Mile High 2A	16

Track on Scene Analysis and Understanding

6th Workshop on 3D Scene Understanding for Vision, Graphics, and Robotics ...	6/4/2026 ... AM	610/612	30
OpenSUN3D: 6th Workshop on Open-World 3D Scene Understanding with Foundation Models	6/3/2026 ... PM	705/707	25
Pixel-level Video Understanding in the Wild Challenge	6/4/2026 ... PM	502	44

Track on Transparency, Safety, Fairness, Accountability, and Ethics in Vision

Machine Unlearning for Vision	6/3/2026 ... PM	Mile High 1AB	25
The 3rd Workshop on New Trends in AI-Generated Media and Security	6/4/2026 ... AM	102/104	30
Trustworthy, Robust, Uncertainty-Aware, and Explainable Visual Intelligence and Beyond	6/4/2026 ... AM	705/707	34

Track on Video: Action and Event Understanding

6th International Workshop on Long-form Video Understanding, Generation and Action	6/4/2026 ... PM	704/706	43
8th International Workshop on Large Scale Holistic Video Understanding	6/4/2026 ... PM	Mile High 4EF	43
Computer Vision with Small Data: Beyond Scale -- Toward Data-Efficient Dynamically-Aware Video Intelligence	6/3/2026 ... PM	102/104	22

Second Workshop on Skilled Activity Understanding, Assessment & Feedback Generation 6/4/2026... PM..... 705/707 44

The 1st Workshop on Vision for Intelligent Task Assistants 6/3/2026... PM..... 108 27

Third Joint Egocentric Vision (EgoVis) Workshop 6/3/2026... Full Day..... 704/706 19

Track on Vision and Graphics

AI for Content Creation 6/3/2026... AM..... 610/612 12

Appearance Understanding and Generation 6/4/2026... PM..... Mile High 4AB 41

See the World in a Different Light: Physical Appearance Modeling and Relighting in the Age of Generative AI 6/3/2026... AM..... 210/212 40

The 3rd AI for Visual Arts Workshop and Challenges 6/3/2026... AM..... Mile High 4AB 12

The 3rd Workshop on AI for Content Generation, Quality Enhancement and Streaming 6/3/2026... AM..... 105 12

Track on Vision Applications and Systems

12th IEEE International Workshop on Computer Vision in Sports 6/4/2026... Full Day..... 503 36

6th Omnidirectional Computer Vision Workshop 6/4/2026... AM..... 711 32

AI4RWC: The 2nd International Workshop on Vision Intelligence for Real-world Challenges 6/3/2026... AM..... 507 21

Artificial Intelligence for Space 6/4/2026... PM..... 710 41

Computer Vision for the Built World 6/3/2026... PM..... 109 22

The 3rd MetaFood Workshop (MTF) 6/3/2026... PM..... 709 25

The 7th International Workshop and CVML Challenge on Agriculture-Vision: Challenges & Opportunities for Computer Vision in Agriculture 6/3/2026... PM..... Mile High 2A 27

Track on Vision for Accessibility

2nd Workshop on Multimodal Sign Language Recognition 6/3/2026... AM..... 103 25

Generative AI for Sign Language 6/3/2026... AM..... 112 13

VizWiz Grand Challenge: Interpreting Images and Videos Taken by Blind People 6/4/2026... AM..... 709 35

Track on Vision for Societal Good

6th Workshop on CV4Animals: Computer Vision for Animal Behavior Tracking and Modeling 6/4/2026... AM..... 108 31

Authenticity & Provenance in the age of Generative AI 6/3/2026... PM..... 103 22

Computer Vision x Education: Building a Cross-Community Agenda for Multimodal Vision in Classrooms 6/4/2026... PM..... 113 42

From Perception to Persuasion: Challenges and Advances in Misinformation Detection in Society 6/4/2026... AM..... 113 33

Track on Vision, Language and Reasoning

2nd Workshop on Knowledge-Intensive Multimodal Reasoning 6/4/2026... AM..... 704/706 32

Cognitive Foundations for Multimodal Models 6/3/2026... PM..... 610/612 22

GRAIL-V: Grounded Retrieval & Agentic Intelligence for Vision-Language 6/3/2026... AM..... 506 14

Multimodal Algorithmic Reasoning Workshop 6/4/2026... AM..... 601 32

The 2nd Workshop on Multi-Modal Reasoning for AI Agents 6/4/2026... PM..... Mile High 3A 44

The 2nd Workshop on Test-time Scaling for Computer Vision 6/3/2026... PM..... 507 27

Workshop on Vision-based Assistants in the Real-World 6/3/2026... AM..... 102/104 17

Workshop on Visual Concepts 6/4/2026... Full Day..... 501 39

Track on World Models

4D World Models: Bridging Generation and Reconstruction 6/4/2026... PM..... 203 44

End-to-End 3D Learning 6/3/2026... PM..... 501 23

Geometry-Free Novel View Synthesis and Controllable Video Models 6/4/2026... AM..... 607 31

The Seventh Annual Embodied Artificial Intelligence Workshop 6/4/2026... Full Day..... 107 36

Tuesday, June 2

14:00 - 20:00 **Registration / Badge Pickup**, Lobby A

Wednesday, June 3

NOTE: Tutorial rooms are subject to change. Refer to the online site for up-to-date locations. Use the QR code for each tutorial to see its schedule. Here is the QR code for the CVPR 2026 Tutorials page.



7:00 - 17:00 **Registration / Badge Pickup**, Lobby A

7:00 - 17:00 **Mother's Room**, Adjacent to Room 102

7:00 - 17:00 **Quiet Room**, Adjacent to Room 102

7:00 - 17:00 **Prayer Room**, Room 206

7:00 - 9:00 **Breakfast**, ExHall C

7:30 - 15:00 **Poster Pickup**, ExHall A

8:00 - 18:00 **WORKSHOPS / TUTORIALS**

8:00 - 18:00 **Posters**, ExHall A

10:00 - 11:00 **Coffee Break**, ExHall A

12:00 - 13:45 **Lunch**, ExHall C

15:00 - 16:00 **Coffee Break**, ExHall A

TUTORIALS

The Principles of Diffusion Models: From Origins to Real-Time Diffusion & Tokenized Vision Models

Organizers: Tsui-Wei (Lily) Weng, Tuomas Oikarinen, Ge Yan, Akshay Kulkarni

Date: 6/3/2026

Time: AM

Location: 301/302



Summary: In recent years, diffusion models have become a central paradigm in computer vision, powering advances in image synthesis, editing, and video generation. However, existing tutorials are often fragmented, focusing either on specific applications or isolated methodological perspectives without a unifying framework. This tutorial aims to present a principle-driven view of diffusion models by distilling their foundations into a small set of core ideas that unify variational, score-based, and flow-based approaches. It further emphasizes emerging directions in real-time generation through flow-map models, which enable fast and interactive visual applications. In addition, the tutorial extends this framework to discrete and tokenized diffusion models, highlighting their role in bridging continuous vision generation with multimodal and structured representations.

From Perception to Simulation: The Emergence of World Models in Multi-modal Reasoning

Organizers: Yujun Cai, Jianfei Cai, Yiwei Wang, Ming-Hsuan Yang

Date: 6/3/2026

Time: PM

Location: 301/302



Summary: World models are emerging as a new paradigm in computer vision and multimodal learning, enabling systems to move beyond perception toward reasoning, simulation, and decision-making. This tutorial explores how world models have evolved from predictive frameworks into engines for multi-modal reasoning, capable of simulating environments, supporting counterfactual thinking, and enabling planning. It examines key approaches for learning world dynamics from visual data, including both discrete tokenization and diffusion-based methods, and highlights their role in modeling physical and causal structure. The tutorial further covers how these models support reasoning through simulation, as well as their applications in embodied agents and robotics, while discussing key challenges such as grounding, scalability, and causal understanding.

Principled Interpretability in Vision Models: From Mechanistic Understanding to Interpretable Models by Design

Organizers: Tsui-Wei (Lily) Weng, Tuomas Oikarinen

Date: 6/3/2026

Time: PM

Location: Mile High 3C



Summary: As deep learning systems are increasingly deployed in high-stakes applications, understanding their internal behavior is essential for ensuring trust, safety, and reliability. However, the field of interpretability remains fragmented, spanning diverse methods without a unified framework or standardized evaluation. This tutorial aims to provide a comprehensive overview of interpretability in vision models, bridging post-hoc mechanistic analysis with approaches that design inherently interpretable models. It reviews techniques for analyzing neural networks at multiple levels—from individual neurons to circuits—alongside recent advances in evaluating the faithfulness of explanations. In addition, the tutorial covers emerging methods for learning interpretable models by design, such as concept-based approaches, and highlights practical applications in debugging, model editing, and safety auditing.

Towards Safe Multi-Modal Learning: Evolving Threats and Safety Solutions

Organizers: Xi Li, Manling Li, Muchao Ye

Date: 6/3/2026

Time: AM

Location: Mile High 3C



Summary: Multi-modal learning has enabled powerful systems that combine text, images, audio, and video for perception, reasoning, and decision-making. At the same time, it has introduced safety challenges that differ fundamentally from those in traditional uni-modal learning. This tutorial presents a structured overview of the evolving safety landscape in multi-modal AI, focusing on emerging threat models and corresponding defense strategies. It examines key risks such as compromised modality integration, modality misalignment, and fused cross-modal vulnerabilities, and reviews recent work on adversarial attacks, jailbreaks, hallucinations, and safety solutions for more reliable multi-modal systems.

Edge AI in Action: Mastering On-Device Inference

Organizers: Fabricio Batista Narcizo,
Elizabeth Munzlinger,
Sai Narsi Reddy Donthi Reddy,
Shan Ahmed Shaffi

Date: 6/3/2026

Time: AM

Location: 702



Summary: Edge AI enables real-time, low-latency inference directly on devices, but achieving high performance and efficiency requires specialized optimization and deployment techniques tailored to heterogeneous hardware. This tutorial provides a hands-on guide to on-device inference, focusing on end-to-end workflows for optimizing and deploying deep learning models on leading edge platforms such as Qualcomm Snapdragon and NVIDIA Jetson. It covers key techniques including model compression, quantization, and hardware-aware optimization, along with practical tools such as SNPE and TensorRT. Through comparative analysis and real-world case studies, the tutorial highlights best practices for achieving efficient, low-latency performance in applications ranging from computer vision to multimodal AI.

Monte Carlo physical simulation

Organizers: Rohan Sawhney, Bailey Miller,
Ioannis Gkioulekas, Keenan Crane

Date: 6/3/2026

Time: PM

Location: 702



Summary: Partial differential equations (PDEs) play a central role in physics-based modeling across vision, graphics, and robotics, but conventional grid-based solvers often struggle with scalability and complex geometry. This tutorial introduces grid-free Monte Carlo methods for solving PDEs, focusing on algorithms such as walk on spheres and walk on stars that eliminate the need for spatial discretization. It presents the theoretical foundations of these methods alongside practical techniques for efficient sampling, variance reduction, and differentiable simulation. The tutorial also highlights applications in vision and robotics, including inverse problems and physics-based learning, and provides hands-on guidance for implementing Monte Carlo PDE solvers in real-world systems.

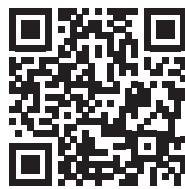
Accelerated Diffusion Models: From Theory to Interactive World Models

Organizers: Julius Berner, Weili Nie,
Arash Vahdat

Date: 6/3/2026

Time: AM

Location: 201



Summary: Diffusion models have become a cornerstone of modern generative modeling, but their practical deployment in interactive applications is often limited by slow and computationally expensive sampling. This tutorial focuses on recent advances in accelerating diffusion models, providing a comprehensive overview of methods that enable fast and efficient generation. It covers general acceleration techniques, training-based approaches such as distillation into few-step samplers, and practical strategies for scaling to image and video generation. The tutorial further highlights how these advances enable emerging applications such as interactive world models and real-time generative systems, and provides hands-on guidance through the FastGen library.

Building GenAI based Simulation Environment for End-to-End Autonomous Driving

Organizers: Henry Liu, Howie Sun, Jun Gao,
Shuo Feng, Xintao Yan, Jiawei Wang

Date: 6/3/2026

Time: PM

Location: 201



Summary: Generative AI is transforming simulation for autonomous driving, enabling data-driven and closed-loop environments that better capture the complexity of real-world scenarios. This tutorial presents an end-to-end framework for building generative simulation pipelines tailored to modern learning-based driving systems. It covers key components including world modeling and city-scale digital twins, generative synthesis of rare and safety-critical scenarios, and realistic sensor and video simulation using both graphics and neural approaches. The tutorial further discusses system-level evaluation and integration with autonomous driving stacks, providing practical guidance and open-source tools for developing scalable and reliable simulation environments.

Tom Builds, Tom Breaks: Hands-On Attacks and Defenses for Vision-Language Systems

Organizers: Pavan Reddy

Date: 6/3/2026

Time: AM

Location: Mile High 2B



Summary: Vision-language models are increasingly deployed in real-world systems where images can directly influence decisions and actions, creating new security risks beyond traditional text-based attacks. This tutorial provides a hands-on introduction to attacks and defenses for vision-language systems, using a practical, end-to-end workflow that mirrors real deployment scenarios. It covers a range of vulnerabilities, including visual jailbreaks, preprocessing-induced attacks, adversarial perturbations, backdoored models, and data poisoning, along with corresponding mitigation strategies. Through interactive examples and reproducible notebooks, the tutorial emphasizes how these threats manifest in practice and how to build robust, auditable systems for multimodal AI.

3D Human Mesh Modeling and Recovery from RGB and LiDAR

Organizers: Romain Brégier, Salma Galaoui,
István Sárádi, Fabien Baradel,
Nermin Samet, David Picard

Date: 6/3/2026

Time: PM

Location: Mile High 2B



Summary: Understanding human pose and shape through parametric body models is a key enabler of applications from AR/VR and sports analysis to human-robot interaction. This tutorial provides an in-depth overview of parametric body models and their role in Human Mesh Recovery. We cover fundamental principles and recent developments, guiding practitioners through major models (e.g., SMPL, Anny, MHR, SOMA) and their trade-offs. We then present state-of-the-art Human Mesh Recovery methods, with a focus on challenging in-the-wild settings across different input modalities, including single- and multi-view RGB, video, depth and LiDAR.

Analytic understanding of diffusion models

Organizers: Artem Lukoianov, Chenyang Yuan,
Christopher Scarvelis,
Mason Kamb

Date: 6/3/2026

Time: Full day

Location: Mile High 3C



Summary: Diffusion models achieve state-of-the-art performance in generative modeling, yet their theoretical foundations and generalization behavior remain poorly understood. This tutorial focuses on the analytical understanding of diffusion models, addressing the apparent paradox between closed-form optimal denoisers and the empirical success of deep diffusion networks. It introduces recent theoretical advances that explain how mechanisms such as score smoothing, training dynamics, neural network inductive biases, and data structure contribute to generalization. By combining mathematical insights with hands-on experiments, the tutorial provides a principled framework for understanding the inner workings of diffusion models and for interpreting recent developments in the field.

WORKSHOPS

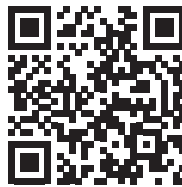
AERO-HPR: Human Perception and Recognition in Aerial Surveillance

Organizers: Arnold Wiliem, Arun Ross,
Clinton Fookes, Dana Michalski,
Hugo Proenca, Huy Nguyen,
Kien Nguyen, Vitomir Struc,
Xiaoming Liu

Date: 6/3/2026

Time: AM

Location: 110



Summary: Aerial person recognition has emerged as a critical research area at the intersection of computer vision and biometrics, driven by commercial UAV accessibility and applications ranging from search-and-rescue to public safety. However, state-of-the-art ground-based methods experience substantial performance degradation when applied to aerial data due to extreme viewing angles, atmospheric turbulence, altitude-induced resolution gaps, and motion blur. This workshop provides the first dedicated forum addressing the complete aerial person analysis pipeline—including detection, tracking, recognition, re-identification, action recognition, and group analysis—with explicit focus on aerial-specific technical challenges. We bring together researchers from computer vision, biometrics, and surveillance communities to advance novel methods and foster responsible development. Community deliverables include consensus statements on ethical research practices, standardized evaluation protocols, and a responsible research checklist for this rapidly growing field.

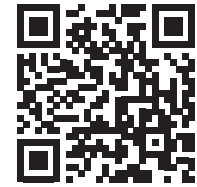
AI for Content Creation

Organizers: James Tompkin, Jun-Yan Zhu,
Lingjie Liu, Yiqing Liang, Yuheng Li,
Thao Nguyen, Krishna Kumar Singh

Date: 6/3/2026

Time: AM

Location: 610/612



Summary: Content creation is vital in photography, gaming, and design. AI and machine learning are revolutionizing this work, with generative models automating tasks to produce photorealistic 2D/3D content, video, and vector from text. These AI techniques also create diverse synthetic data for downstream tasks. The “AI for Content Creation” workshop unites world-class experts and researchers to explore this exciting, fast-moving field. Now in its 7th year, the workshop fosters learning, discussion, and networking for this popular research area. <https://ai-for-content-creation.github.io/>

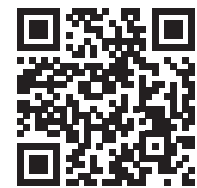
The 3rd AI for Visual Arts Workshop and Challenges

Organizers: Deblina Bhattacharjee,
Iris (Yin) Zhang, Bingchen Zhao,
Haoxiang Li, Luoqi Liu

Date: 6/3/2026

Time: AM

Location: Mile High 4AB



Summary: The 3rd AI for Visual Arts (AI4VA) workshop at CVPR 2026 examines how computer vision perceives, interprets, and models abstraction across artistic and cultural imagery like paintings, comics, sculptures, and installations that challenge conventional models. These stylized domains expose weaknesses in robustness, generalization, and interpretability, offering a principled framework for evaluating perception beyond realism. Can models recognize the “dog-ness” in a Picasso, or infer motion from comic lines? Building on ECCV 2024 and ICCV 2025 editions, AI4VA at CVPR 2026 is organized around three pillars: Perception Under Abstraction (multimodal understanding under stylization), Provenance & Authenticity (authenticity detection and tracing AI-assisted creative transformations), and Cross-Disciplinary Exchange between artists, vision researchers, and curators. The workshop also hosts the PortraitCraft Challenge (co-organized with Meitu) with two tracks: Track 1 - Portrait Composition Understanding (scoring, fine-grained attributes, VQA), and Track 2 - Portrait Composition Generation (from structured composition descriptions).

The 3rd Workshop on AI for Content Generation, Quality Enhancement and Streaming

Organizers: Radu Timofte, Marcos V. Conde

Date: 6/3/2026

Time: AM

Location: 105



Summary: Streaming is a \$50.11 billion industry where hundreds of millions of users demand everyday high-quality content on different platforms. Nowadays, most of the content is enhanced or generated using AI. Computer vision and deep learning allow the development of novel methods for rendering and generating content, compression, enhancement, and quality assessment. From neural codecs to deep learning-based video enhancement and on-device generation. Continuing with the success of previous editions, we propose the The 3rd Workshop on AI for Content Generation, Quality Enhancement and Streaming, the largest workshop focused on unifying content generation and streaming.

Workshop on “Bitter Lessons”

Organizers: Aditya Prakash, Anand Bhattad, Angjoo Kanazawa, Dima Damen, Georgia Gkioxari, Svetlana Lazebnik, Unnat Jain

Date: 6/3/2026

Time: AM

Location: Four Seasons 4



Summary: Between deadlines, bigger lessons slip through the cracks, buried in rejected papers, abandoned projects, or overlooked work. This workshop explores the “bitter lessons” of computer vision: hard-won wisdom our field has accumulated but rarely discusses. We’ll examine Pyrrhic victories that won benchmarks but lost larger battles, “failed” ideas that were simply ahead of their time, concepts that keep resurfacing, and cases where simple methods unexpectedly triumphed. By openly reflecting on what worked, what didn’t, and why, we aim to learn from our collective experience, reducing repeated mistakes and distilling insights to guide the next generation of research.

AI for Creative Visual Content Generation, Editing and Understanding

Organizers: Alex Schwing, Dong Liu, Jay Maanav Mahajan, Jiaju Ma, Junho Kim, Maneesh Agrawala, Ozgur Kara, Rushikesh Zawar, Victor Escorcía, Fabian Caba Heilbron, James Matthew Rehg

Date: 6/3/2026

Time: AM

Location: 501



Summary: The 9th CVEU workshop bridges generative AI research with artistic creation. We bring together researchers, artists, and entrepreneurs to explore the next generation of AI-driven creative tools. This year, we introduce a special theme on “world modeling for creative video applications,” focusing on generating consistent and dynamic scenes crucial for ‘interactive’ storytelling and video synthesis. Through keynotes, a panel discussion, and submitted papers, we aim to foster collaboration and explore how GenAI is reshaping creative practices, from visual production to next-generation interactive media.

The 5th DataCV Workshop and Challenge

Organizers: Hazel Doughty, Jose Lezama, Kun Yuan, Liang Zheng, Xiaoxiao Sun, Yue Yao, Yuhui Zhang, Manmohan Chandraker, Serena Yeung-Levy

Date: 6/3/2026

Time: AM

Location: 710



Summary: DataCV workshop advances data-centric perspectives in vision and multimodal AI, shifting focus from algorithms to the design, analysis, and optimized use of datasets to better understand and improve models. We explore dataset-level properties, representations, and similarities, and challenges in bias, fairness, leakage, and generalization, with a focus on robust model evaluation through effective dataset design. Topics include evaluating vision-language models, improving dataset quality through simulation, and reducing reliance on labeled data. The workshop encourages research on how dataset insights can guide model development, performance prediction, and ethical considerations. By fostering discussion and innovation in vision data, DataCV promotes more robust, generalizable, and responsible vision capabilities in AI systems.

The 22nd Embedded Vision Workshop

Organizers: Ahmed Nabil Belbachir, Branislav Kisacanin, Matteo Poggi, Tse-Wei Chen

Date: 6/3/2026

Time: AM

Location: 709



Summary: Embedded vision is an active field of research, bringing together efficient learning models with fast computer vision and pattern recognition algorithms, to tackle many areas of robotics and intelligent systems that are enjoying an impressive growth today. Such strong impact comes with many challenges that stem from the difficulty of understanding complex visual scenes under the tight computational constraints required by real-time solutions on embedded devices. The Embedded Vision Workshop will provide a venue for discussing these challenges by bringing together researchers and practitioners from the different fields outlined above.

The 5th Workshop on Federated Learning for Computer Vision

Organizers: Aritra Dutta, Chen Chen, Guangyu Sun, Mahdi Morafah, Nathalie Baracaldo, Victor Zhu, Yang Liu, Zhishuai Guo, Nicholas D. Lane

Date: 6/3/2026

Time: AM

Location: 711



Summary: The growing shift from centralized clouds to edge devices is reshaping AI. Federated Learning (FL) has emerged as a key technology enabling large-scale, privacy-preserving intelligence at the edge. For computer vision—where data are rich in semantics and privacy-sensitive—FL offers unique opportunities and challenges in achieving fairness, efficiency, and trustworthiness across distributed data sources. Building on four successful editions at CVPR 2022–2025, FedVision-2026 expands its focus to foundation-model adaptation, personalized and efficient edge learning, and trustworthy visual intelligence, fostering collaboration across academia, industry, and open-source communities to define the next frontier of distributed visual learning.

Generative AI for Sign Language

Organizers: Hezhen Hu, Yuecong Min, Ronglai Zuo, Oscar Koller, Léore Bensabath, Wengang Zhou, Houqiang Li, Stefanos Zafeiriou, Xilin Chen, Hongdong Li, and Dimitris N. Metaxas

Date: 6/3/2026

Time: AM

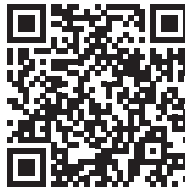
Location: 112



Summary: Sign language, as a rich and structured visual language, encodes linguistic meaning through fine-grained gestures, facial expressions, and body movements. Recent advances in sequential generative models and diffusion-based approaches offer a promising pathway to produce high-quality translations, synthesize realistic and expressive digital signers, expand low-resource sign datasets, and foster more inclusive communication between deaf and hearing communities. This workshop brings together researchers from computer vision, natural language processing, linguistics, and accessibility studies to explore the frontiers of generative modeling for sign language and to foster responsible, human-centered AI systems that understand and communicate through this uniquely visual language.

Generative AI for XR and Identity-based Applications

Organizers: Brendan David-John, Emanuela Marasco, Long Chen, Pinar Yanardag, Shengwei An, Chris Thomas



Date: 6/3/2026

Time: AM

Location: 111

Summary: The convergence of Generative AI, Extended Reality (XR), and biometric technologies is reshaping the landscape of XR sensing and authentication systems. The workshop will feature speakers, paper presentations, and introduce a multimodal XR dataset to seed future research competitions. We invite contributions relevant to at least two of our three topic domains through novel architectures, experiments, and frameworks. By bringing together experts across computer vision, AI, and XR GenXR-ID aims to chart a path toward secure XR systems powered by generative AI.

GigaBrain Challenge 2026: Workshop on World Models Empowering Vision Language Action Model

Organizers: Zheng Zhu, Xiaofeng Wang, Hongyang Li, Shanghang Zhang, Yao Mu, Haoqiang Fan, Zhizhong Su



Date: 6/3/2026

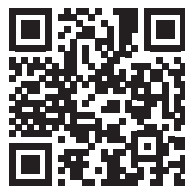
Time: PM

Location: 507

Summary: This GigaBrain Challenge 2026 Workshop advances embodied intelligence by connecting Vision-Language-Action (VLA) and World Models. It features paper presentations, poster sessions, and the GigaBrain Challenge across four tracks: RoboTwin (simulation), GigaWorld (World Model as evaluator), RoboChallenge (real-robot), and PhysClaw* (creative OpenClaw applications). Papers will be selected with a Best Paper Award. The workshop will host award ceremonies and winning-team presentations. We aim to build benchmarks, foster collaboration, and drive real-world generalization for embodied AI.

GRAIL-V: Grounded Retrieval & Agentic Intelligence for Vision-Language

Organizers: Amit Agarwal, Vivek Gupta, Vivek Srikumar, Tao Sheng, Alice Oh, Sara Hooker, Jyotika Singh, Hitesh Laxmichand Patel



Date: 6/3/2026

Time: AM

Location: 506

Summary: GRAIL-V is a CVPR workshop dedicated to grounded multimodal retrieval and reranking for agentic-vision systems. As agents plan, retrieve, and verify before acting across diverse structured and unstructured data sources, robust evidence grounding and efficiency become bottlenecks for reliable deployment. GRAIL-V brings together CV, IR, NLP, and HCI communities to advance unified methods and evaluation practices for visual-text search, multimodal tool use, and calibrated decision-making. GRAIL-V features keynotes, presentations, and a panel, emphasizing real-world constraints. By fostering shared evaluation protocols and cross-domain collaboration, GRAIL-V aims to catalyze progress toward robust, accountable, and deployable agentic-vision systems, engaging both researchers and practitioners.

The 3rd Workshop on Human Motion Generation - New Perspective on Simulation, Animation, and VR applications

Organizers: Christian Theobalt, Guy Tevet, Ikhsanul Habibie, Peizhuo Li, Rishabh Dabral, Sigal Raab, Yuxuan Mu, Chuan Guo, Amit H. Bermano



Date: 6/3/2026

Time: AM

Location: 505

Summary: Human motion is a fundamental attribute of life, shaping actions, behavior, and social interaction. Recent advances in generative motion modeling have experienced rapid progress across CVPR-relevant areas—including embodied intelligence, robotics learning, 3D character animation, AR/VR, crowd simulation, and sports analytics—by translating goals and prompts into controllable, physically plausible movements. This third edition of the Workshop on Human Motion Generation (HuMoGen) foregrounds new perspectives on **simulation, animation, and VR applications**. The program will feature invited talks from leading experts, contributed papers, covering topics on physical simulation, interaction, controllability, dataset, scalable generation. Building on two successful CVPR editions (2024–2025) with 200+ attendees annually, HuMoGen aims to catalyze cross-pollination between vision, graphics, simulation, and XR communities, shaping the next generation of methods and best practices for responsible, scalable human motion generation.

IPA: Interactive Physical AI Workshop

Organizers: Amrita Mazumdar, Koki Nagano, Leena Mathur, Seonwook Park, Shalini De Mello, Shengze Wang



Date: 6/3/2026

Time: AM

Location: 203

Summary: The 1st Interactive Physical AI (IPA) Workshop (held in conjunction with CVPR 2026) brings together researchers from computer vision, robotics, and multi-modal AI to explore how intelligent systems can perceive, communicate, and act safely and naturally alongside humans in the shared physical world. Physical AI spans embodied agents such as humanoid robots, physically-grounded and environment-aware avatars, and on-device audio-visual assistants capable of real-time perception and social communication. The workshop addresses key challenges in cooperative human-AI tasks, egocentric multimodal perception, naturalistic audio-visual generation, safe navigation, personalization across users and environments, privacy-aware and lifelong learning, and low-latency inference for interactive systems. The IPA 2026 workshop homepage can be found at: <https://research.nvidia.com/labs/amri/projects/IPA/2026/>

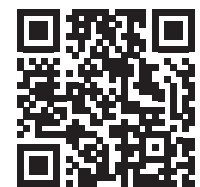
LatinX in Computer Vision Research Workshop

Organizers: Ana María Quintero-Ossa

Date: 6/3/2026

Time: AM

Location: 106



Summary: This workshop, a key initiative of LatinX in AI, directly addresses the systemic underrepresentation of the LatinX community in Computer Vision (CV) to ensure equitable access and ethical AI development. The event serves as a vital platform for LatinX researchers at all levels to discuss academic, industrial, and social challenges. We will boost visibility by highlighting prominent LatinX researchers and allies, with a special focus on LatinX women. Furthermore, we drive career growth through sponsored registrations, struc-

tured mentoring, and resume sharing. By increasing LatinX participation at major conferences, we foster a more diverse and innovative CV ecosystem. While focused on LatinX researchers, all are invited to attend.

Multimodal Alignment for a Pluralistic Society

Organizers: Aishwarya Agrawal, Aniket Rege, Danish Pruthi, Desmond Elliott, Karolina Stanczak, Mehar Bhatia, Negar Rostamzadeh, Shravan Nayak, Simran Khanuja, Yash Goyal



Date: 6/3/2026
Time: AM
Location: 113

Summary: Multimodal AI systems are fundamentally reshaping how we create, communicate, and access information globally. However, they often represent a narrow slice of society as they lack sufficient cultural, geographic, and linguistic awareness. This limitation causes models to misrepresent cultural practices, reinforce stereotypes, and promote a homogenized worldview, creating a fundamental misalignment with our diverse, pluralistic world. This workshop brings together interdisciplinary researchers from computer vision, NLP, HCI, social sciences, and humanities to advance the alignment of multimodal AI systems with diverse cultures through invited talks, paper presentations, and the MTV challenge, ensuring they are truly representative of our global society.

Multimodal Foundation Models for Biomedicine: Challenges and Opportunities

Organizers: Dongxia Wu, Hoifung Poon, Jeffrey K Jopling, Xiaohan Wang, Yuhui Zhang, Yuzhe Yang, Serena Yeung-Levy, Yanran Wang

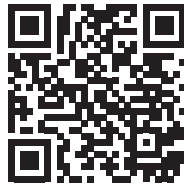


Date: 6/3/2026
Time: AM
Location: Mile High 1CD

Summary: Biomedical data spans diverse modalities—from molecular genomics and cellular microscopy to pathology, radiology, and electronic health records. This workshop unites experts to address key questions: (1) Technical Challenges: What are the core limitations of current multimodal learning in biomedicine? Issues include aligning data across spatial and temporal scales, handling annotation imbalances, and preserving modality-specific contexts while enabling cross-domain transfer. (2) New Opportunities: What breakthroughs can multimodal foundation models enable? Promising directions include multi-scale disease diagnosis, personalized treatment integrating wearables and genomics, and context-aware clinical operations linking surgical videos with patient records.

The Second CVPR Workshop on Foundation and Large Vision Models in Remote Sensing (MORSE)

Organizers: Saurabh Prasad, Jocelyn Chanussot, Begüm Demir, Biplab Banerjee, Danfeng Hong



Date: 6/3/2026
Time: AM
Location: 106

Summary: CVPR MORSE brings together the computer vision, machine learning, and remote sensing communities to explore foundation models, large vision models, vision-language models, and related emerging approaches for Earth observation. The workshop focuses on how recent advances in modern vision models can support remote sensing prob-

lems involving multi-sensor, multi-modal, multi-temporal, and multi-scale data, including optical, hyperspectral, SAR, and LiDAR imagery. MORSE welcomes both methodological advances and application-driven contributions, spanning model training, adaptation strategies, prompt and instruction tuning, discriminative and generative modeling, few-shot and continual learning, open-set recognition, and downstream tasks such as segmentation, classification, regression, object detection, and change detection. The workshop features invited keynotes, peer-reviewed papers, and extended abstracts presenting late-breaking results. Through these presentations and discussions, MORSE aims to highlight opportunities and challenges in applying modern vision models to geospatial image analysis, with applications in environmental monitoring, climate science, urban analysis, and broader Earth observation.

The 2nd Workshop on Multimodal Spatial Intelligence

Organizers: Juil Koo, Leonidas Guibas, Mikaela Angelina Uy, Minhuk Sung, Sanja Fidler, Songyou Peng, Phillip Y. Lee



Date: 6/3/2026
Time: AM
Location: 601

Summary: The Multimodal Spatial Intelligence (MUSI) workshop explores how multimodal large language models (LLMs) understand, reason about, and interact with spatial information from the physical world. Spatial intelligence inherently involves integrating images, videos, and 3D data, requiring collaboration across computer vision, robotics, graphics, and NLP. While recent LLMs exhibit strong visual-spatial reasoning, key challenges remain – including understanding spatial relationships, 3D environment modeling, and real-world spatial reasoning. This workshop investigates how LLMs learn cross-modal spatial representations, enhance world modeling and embodied AI, and incorporate ethical and responsible AI considerations. Our goal is to establish benchmarks, promote cross-disciplinary collaboration, and accelerate progress toward robust spatial reasoning in multimodal AI systems.

On Sensor Vision Workshop

Organizers: Laurie Bose, Maciej Lewandowski, Piotr Dudek, Shinjeong Kim, Walterio W. Mayol-Cuevas,



Andrew J. Davison
Date: 6/3/2026
Time: AM
Location: Mile High 4EF

Summary: Many exciting computer vision applications operate at the edge; this could be in drones, IoT devices, or AR/VR systems. These demand visual processing that is fast, energy-efficient, and privacy-preserving. On-sensor vision meets these needs by unifying sensing and computation within a single chip, producing information-rich outputs instead of raw images. By minimizing costly data transfers, such architectures enable real-time, low-power operation. Prototypes like Manchester's SCAMP integrate sensing, memory, and parallel processing directly on-chip. This workshop explores algorithms and architectures for near-pixel, in-sensor, and stacked sensor-processor devices, and how to best partition computation between on-sensor and off-sensor processing.

2nd Workshop on Photorealistic 3D Head Avatars

Organizers: Justus Thies, Koki Nagano, Shalini De Mello, Simon Gehlbach, Tianye Li, Tobias Kirschstein, Matthias Nießner

Date: 6/3/2026

Time: AM

Location: 107



Summary: Photorealistic 3D head avatars will play a crucial role in future computer games, visual effects, movie production, and virtual telepresence. In this workshop, we bring together leading academic researchers and industry experts to discuss the technology behind 3D head avatars, current applications, and future trends. In particular, the workshop hosts a challenge on the NeRSemble 3D Head Avatar benchmark. Participants are invited to submit their methods for two tasks: FLAME-driven Monocular 3D Head Avatar Creation and 3D Face Tracking. The authors of the best-performing submission will receive a GPU prize and present their method alongside invited speakers in the workshop.

22nd Workshop on Perception Beyond the Visible Spectrum

Organizers: Riad I. Hammoud, Yi Ding, Michael Teutsch, Angel D. Sappa, Olaf Wysocki, Robert Laganière, Oliver Nina, Wassim El Ahmar, Leo Ramos

Date: 6/3/2026

Time: AM

Location: 712



Summary: The Perception Beyond the Visible Spectrum workshop series (IEEE PBVS), established over 2 decades ago in 2004, has consistently been a key event within the computer vision and pattern recognition (CVPR) community. This prestigious series has continually highlighted cutting-edge sensors, sensing technologies, and exploitation algorithms operating in the non-visible spectrum, including infrared, thermal, radar, Radio Frequency (RF), hyper-spectral, sonar, lidar, acoustic, and x-ray imaging/sensing. It has been a leading venue for students, engineers and scientists from academia, industry, and government agencies throughout the world.

The 2nd International Workshop & Challenge on Subtle Visual Computing @CVPR 2026

Organizers: Adam Czajka, Adams Wai-Kin Kong, Björn Schuller, Dan Guo, Ghada Khoriba, Philip Torr, Qi Dou, Sergio Escalera, Xiaobao Guo, Xiaochun Cao, Xun Lin, Zitong YU, Shuo Ye, Yefeng Zheng

Date: 6/3/2026

Time: AM

Location: 109



Summary: Subtle visual signals, though often imperceptible to the human eye, contain subtle yet crucial information that can reveal hidden patterns within visual data. To detect and analyze subtle signals has profound implications across various fields, e.g., (1) from biometric and multimedia security, where detecting high-fidelity physical spoof and digital manipulation can enhance the trustworthiness and reliability of real-world visual systems; (2) from medicine, where early identification of minute anomalies in medical imaging can lead to life-saving interventions, (3) from industry,

where spotting micro-defects in production lines can prevent costly failures, (4) from affective computing, where understanding micro-expression, micro-gesture, and hidden physiological signals under human interaction scenarios can benefit the deception detection. This workshop seeks to develop innovative representation learning models specifically designed to capture and interpret subtle visual signals.

1st Workshop on Video World Models: Interaction, Memory, and Efficiency

Organizers: Gordon Wetzstein, Jiajun Wu, Jiwen Yu, Ming-Yu Liu, Philip Torr, Xihui Liu, Yiran Qin, Mike Zheng Shou

Date: 6/3/2026

Time: AM

Location: 705/707



Summary: World models are systems that predicts the future state of the world based on past states and interactive action control. Video World Models, based on interactive video generation, are a promising approach due to their photorealism, scalability, and recent advancements. However, current video generation models are still far from achieving the ideal Video World Model, facing challenges in interaction, memory, efficiency, and downstream applications. This workshop aims to provide a platform for researchers to discuss these challenges and foster collaboration to advance the development and practical application of Video World Models.

Women in Computer Vision

Organizers: Carla Muntean, JungEun Kim, Karen Sanchez, Paula Ramos-Giraldo, Georgia Fargetta, Tatiana Gelvez Barrera, Salimeh Sekeh, Sara EL-ATEF

Date: 6/3/2026

Time: AM

Location: 708



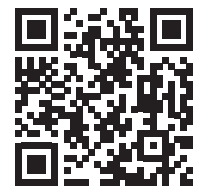
Summary: The Women in Computer Vision (WiCV) Workshop provides a platform to celebrate, connect, and support women researchers in computer vision, addressing the gender gap in the field while fostering mentorship, visibility, and collaboration. Computer vision has become one of the largest research areas in computer science, with major advances in perception, video analysis, 3D reconstruction, generative AI, medical imaging, art and AI, and multimodal large language models. Yet, the percentage of female researchers in academia and industry remains low, leaving many women with limited networking opportunities. WiCV aims to: raise the visibility of women researchers through invited talks by female role models; offer opportunities for junior women to present their work via posters and travel awards; and share experiences and career advice from leaders in the field. This half-day workshop welcomes all researchers in computer vision and seeks to strengthen connections across career levels. Attendance grants will be provided to selected female presenters, supported by generous sponsors each year.

Workshop on World Models Meet Active Sensing and Closed-Loop Planning

Organizers: Alan Yuille, Chen Wei, Cheng Peng, Jianwen Xie, Rama Chellappa, Sanjeev Khudanpur, Tianmin Shu, Yilun Du, Jieneng Chen

Date: 6/3/2026

Time: AM



Location: Mile High 2A

Summary: Recent advances in world models have enabled unprecedented capabilities in image synthesis, video generation, and multimodal simulation. Yet, most remain passive generators—producing outputs from static inputs without active interaction or closed-loop feedback. This workshop bridges generative world models, active sensing, and closed-loop planning, reimagining generative models as active agents that decide what to sense, when to sense it, and how to act. We focus on how generative models can guide information gathering, integrate into model-predictive control, and adapt through continual feedback. By connecting perception, prediction, and action, we aim to uncover how imagination-driven generation can enable intelligent exploration and decision-making in embodied systems. The workshop will also feature a challenge on active embodied tasks within a closed-loop framework, promoting world models that not only generate but also reason, explore, and plan purposefully.

The 5th Explainable AI for Computer Vision (XAI4CV) Workshop

Organizers: Avinab Saha, Dahye Kim, Indu Panigrahi, Jon Donnelly, Lenka Tětková, Marina Gavrilova, Maximilian Dreyer, Miguel-Ángel Fernández-Torres, Sukrut Rao, Yuhui Zhang



Date: 6/3/2026

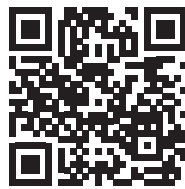
Time: AM

Location: Mile High 1AB

Summary: The Multimodal Spatial Intelligence (MUSI) workshop explores how multimodal large language models (MLLMs) understand, reason about, and interact with spatial information from the physical world. Spatial intelligence inherently involves integrating images, videos, and 3D data, requiring collaboration across computer vision, robotics, graphics, and NLP. While recent MLLMs exhibit strong visual-spatial reasoning, key challenges remain – including understanding spatial relationships, 3D environment modeling, and real-world spatial reasoning. This workshop investigates how MLLMs learn cross-modal spatial representations, enhance world modeling and embodied AI, and incorporate ethical and responsible AI considerations. Our goal is to establish benchmarks, promote cross-disciplinary collaboration, and accelerate progress toward robust spatial reasoning in multimodal AI systems.

Workshop on Vision-based Assistants in the Real-World

Organizers: Angela Yao, Apratim Bhattacharyya, Bugra Tekin, Fadime Sener, Guodong Ding, Rajeev Yasarla, Roland Memisevic, Shweta Mahajan, Leonid Sigal



Date: 6/3/2026

Time: AM

Location: 102/104

Summary: The Second Workshop on Vision-based Assistants in the Real World (VAR) at CVPR 2026 aims to advance research on interactive systems that leverage vision to assist users in everyday tasks such as cooking, fitness, driving, among others. Vision-based assistants promise transformative impact by providing guidance and support in real-world applications. Building on the success of the VAR workshop at CVPR 2025, this workshop will provide a forum for discussing the challenges of accuracy, robustness, privacy, and security – crucial for deployment under real-world conditions. The program will feature invited talks from leading experts, poster sessions, and live demonstrations showcasing practical implementations. The workshop will also host two challenges focused on

interactive task guidance in fitness and cooking domains. By fostering collaboration between academia and industry, VAR 2026 seeks to accelerate progress toward reliable and robust vision-based assistants that can operate effectively in real-world settings.

The 3rd Workshop on Foundation Models for Medical Vision

Organizers: BO WANG, Jun Ma, Yuyin Zhou, Vishal M. Patel

Date: 6/3/2026

Time: AM

Location: 607



Summary: The advent of foundation models has great potential to transform medical vision. These models, with their ability to leverage vast amounts of data and generalize across diverse tasks, offer groundbreaking opportunities for medical imaging and diagnostics. This workshop will explore the role of foundation models in analyzing complex medical images, ranging from radiology and pathology to ophthalmology and cardiology. We will present cutting-edge techniques for interpreting these images, as well as the challenges of integrating these models into real-world clinical environments. Participants will gain valuable insights into how foundation models can drive advancements in diagnostic accuracy, efficiency, and personalized patient care in the medical domain.

PHAROS AI Factory for Medical Imaging & Healthcare

Organizers: Dimitrios Kollias, Francesco Rundo, Stefanos Kollias, Xujiang Ye

Date: 6/3/2026

Time: AM

Location: 108



Summary: The PHAROS-AIF-MIH Workshop and Competition is a premier forum on transparent, secure and human-centered use of AI in medical imaging and healthcare. Building on PHAROS AI Factory and its healthcare vertical, the workshop brings together researchers and practitioners working on CT, MRI, X-ray, EHR and multimodal biomedical data, with particular emphasis on trustworthy, privacy-preserving and regulation-aware AI. It highlights recent advances in generative AI, multimodal LLMs, VLMs, knowledge distillation, agentic AI, continual learning, uncertainty estimation, drift monitoring, fairness and explainability for healthcare applications. The programme promotes interdisciplinary exchange across AI, computer vision, medical imaging, data science and clinical innovation, while supporting the broader goal of translating advanced AI methods into robust and deployable healthcare solutions. The programme includes keynote talks from leading experts, technical paper presentations and a competition focusing on two important challenges: Multi-Source COVID-19 Detection and Fair Disease Diagnosis. Built on real-world multi-source datasets and fairness-aware evaluation protocols, these challenges emphasize robust generalization across hospitals and equitable performance across demographic groups. Overall, PHAROS-AIF-MIH reflects the growing importance of transparent, secure and high-performance AI systems for medical decision support, while fostering benchmarking, collaboration and innovation across the medical imaging and healthcare community in Europe and beyond for impact.

Workshop on Agentic AI for Visual Media

Organizers: Anyi Rao, Enze Xie, Jian Wang, Jing Shao, Jinjin Gu, Lei Sun, Luc Van Gool, Philip Torr, Zhenfei Yin, Danda Pani Paudel, He Zhang



Date: 6/3/2026

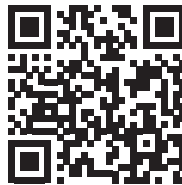
Time: Full Day

Location: Mile High 1EF

Summary: This workshop explores the emerging paradigm of agentic AI for visual media, uniting advances in computer vision, multimodal learning, and intelligent agents. It focuses on autonomous agentic systems that can reason, plan, and orchestrate visual models and tools to accomplish complex media tasks. By bridging research in perception, reasoning, and decision-making, the workshop aims to establish foundational methodologies, foster collaboration between academia and industry, and accelerate progress toward scalable and adaptive visual intelligence systems.

Bridging Vision, Language, and Action: What's Missing in Actionable Visual Perception for Robotics

Organizers: Jiawei Ma, Chengzhi Mao, Peng Wang, Baoru Huang, Anh Nguyen, Zhutian Yang, Kaichun Mo, Nathan F. Lepora, Bolei Zhou, Yu Zhang, Joya Chen, Wenliang Guo, Qinghong Lin



Date: 6/3/2026

Time: Full Day

Location: 503

Summary: Vision foundation models are primarily trained on passive data, encoding the world as pixels or linguistic tokens. Robots, in contrast, must reason about physically-grounded quantities like 6D pose, dynamics, and affordances. This workshop forges a bidirectional co-design channel to translate perceptual abstractions into actionable structures for control and planning. We will explore two key directions: [Vision-to-Robotics] defining novel model outputs linked to robot-centric visual semantics, and [Robotics-to-Vision] learning control-ready representations that directly interface with planning and motor controllers. Our goal is to close the loop from pixels to torque, enabling a new generation of task-hardened, physically-aware AI for robotics.

Autonomous Understanding Through Open-world Perception and Integrated Language models for On-road Tasks

Organizers: Ali K. AlShami, Ryan Rabinowitz, Maged Shoman, Jianwu Fang, Lukáš Pícek, Shao-yuan Lo, Steve Cruz, Lei-lei Li, Nachiket Kamod, Jugal Kalita, Terrance E. BoultonAlShami



Date: 6/3/2026

Time: Full Day

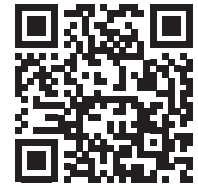
Location: Mile High 3A

Summary: AUTOPILOT is a workshop on safety-critical autonomous driving, spotlighting robust perception and trajectory forecasting that support reliable decision-making and motion planning. It emphasizes the practical use of foundation models, vision-language and generative, through efficient distillation for on-vehicle deployment. A core theme is open-world learning, addressing Out-of-Distribution (OOD) and known hazards by detecting, predicting, and mitigating novel objects, agents, and events be-

yond standard taxonomies. AUTOPILOT features invited talks from leading industry experts, an open challenge, and archival proceedings, bringing academia and practitioners together to develop real-world solutions with explicit attention to societal impact, ethics, and reproducible evaluation.

Computational Cameras and Displays

Organizers: Ayush Bhandari, Fei Xia, Qi Guo, Vishwanath Saragadam



Date: 6/3/2026

Time: Full Day

Location: Mile High 4CD

Summary: This workshop is designed to unite the computational camera and display communities in that it considers to what degree concepts from computational cameras can inform the design of emerging computational displays and vice versa, both focused on applications in computer vision. The Computational Cameras and Displays (CCD) workshop series serves as an annual gathering place for researchers and practitioners who design, build, and use computational cameras, displays, and imaging systems for a wide variety of uses. This year's workshop will have exciting keynote and invited speakers, announcement of a new computer vision challenge, and an exciting poster and demo session. The workshop will host posters and demo submissions on all topics relating to computational imaging systems.

Foundation Models for V2X-Based Cooperative Autonomous Driving

Organizers: Alina Roitberg, Alois Knoll, Arpita Vats, Bolei Zhou, Chuheng Wei, Daniel Watzenig, Guoyuan Wu, Haibao Yu, Holger Caesar, Hu Cao, Jiajie Zhang, Jiaqi Ma, Julie Stephany Berrio Perez, Lars Ullrich, Mao Shan, Max Peter Ronecker, Mohan Trivedi, Peizheng Li, Rahul Raja, Ross Greer, Rui Song, Tianhui Cai, Walter Zimmer, Xiangbo Gao, Xingcheng Zhou, Zewei Zhou, Zhaoliang Zheng, Zhengzhong Tu, Zhiyu Huang, Seth Z. Zhao, Yun Zhang



Date: 6/3/2026

Time: Full Day

Location: 207

Summary: V2X-based cooperative systems can enhance perception, planning, and decision-making capabilities of Autonomous Vehicles (AVs). V2X communication enables vehicles and roadside infrastructure to share real-time information, improving situational awareness, extending sensory range, and providing early warnings for hazards. These cooperative systems address limitations of single-vehicle perception by fusing sensor data from multiple sources, which is crucial for understanding complex traffic scenarios. At the same time, foundation models, such as VLMs, have demonstrated strong generalization abilities by leveraging large-scale, cross-domain data. These models enable zero-shot learning, open-vocabulary object recognition, and scene interpretation, allowing autonomous systems to better handle unseen objects and novel traffic situations. By combining CP with foundation models, AVs can achieve a more robust and adaptable understanding of the environment. Our workshop addresses key challenges and drives innovation toward safer and more efficient autonomous transportation.

Efficient Deep Learning for Computer Vision

Organizers: Chen Chen, Ning Bi, Pavlo Molchanov, Raghuraman Krishnamoorthi, Xiao Hu, Yung-Hsiang Lu, shuai zhang

Date: 6/3/2026

Time: Full Day

Location: 502



Summary: The 9th Efficient Deep Learning for Computer Vision (ECV) Workshop will convene leading researchers and practitioners to advance scalable, resource-aware AI for vision applications. The program features innovations in foundation models, generative AI, 3D scene understanding, embodied agents, and neural rendering for mobile, wearable, and edge devices. Attendees will benefit from invited talks, onsite demos, and the IEEE Low-Power Computer Vision Challenge. ECV provides a premier platform for exchanging ideas and fostering collaboration to shape the future of efficient computer vision on edge devices.

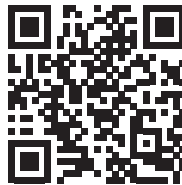
Third Joint Egocentric Vision (EgoVis) Workshop

Organizers: Antonino Furnari, Chiara Plizzari, Dima Damen, Giovanni Maria Farinella, Jitendra Malik, Kristen Grauman, Marc Pollefeys, Masashi Hatano, Richard Newcombe, Siddhant Bansal, Tushar Nagarajan, Yoichi Sato, David J. Crandall

Date: 6/3/2026

Time: Full Day

Location: 704/706



Summary: The Joint Egocentric Vision (EgoVis) Workshop brings together the community advancing first-person perception, covering key advances in embodied AI, video understanding, multi-modal learning, self-supervised learning, and AR/VR applications. Gathering over 100 researchers from academia and industry, EgoVis features keynotes, paper presentations, panels, and challenge results from major datasets including EPIC-KITCHENS, HD-EPIC, Ego4D, Ego-Exo4D, HoloAssist, and EgoCross, fostering collaboration to advance egocentric vision research and its impact on robotics, cognitive science, and everyday technologies.

From Lab Demos to Daily Tasks: Embodied Intelligence in the Wild

Organizers: Hongyang Li, Ping Luo, Hao Zhao, Andrei Bursuc, Animesh Garg, Lingjie Liu, Amy Zhang, Huijie Wang

Date: 6/3/2026

Time: Full Day

Location: Four Seasons 1



Summary: Embodied intelligence brings us closer to creating robots that can truly see, think, and act in the real world. Yet, today's systems remain fragile, struggling to handle the complexity of the real world. This workshop looks ahead - toward generalizable embodied intelligence that thrives beyond the lab. By gathering pioneers from vision and robotics, the workshop seeks to identify key challenges, discuss interdisciplinary solutions, and chart promising directions for advancing embodied systems to everyday life.

13th Workshop on Fine-grained Visual Categorization

Organizers: Aaron Sun, Adriano C. D'Alessandro, Grant Van Horn, Joakim Bruslund Haurum, Laura Chrobak, Lukas Pícek, Madison Van Horn, Nico Lang, Rangel Daroya, Sara Beery, Serge Belongie, Subhansu Maji, Xiangteng He, Oisín Mac Aodha

Date: 6/3/2026

Time: Full Day

Location: 504



Summary: FGVC13 will address topics like self-supervision, limited data, and human-in-the-loop learning, all through the lens of fine-grained learning. Extending beyond traditional computer vision, it highlights real-world applications in ecology, biology, medicine, and the humanities, encouraging participation from diverse research communities. Building on successful challenges like iNaturalist, FGVC13 will feature new competitions. This domain-agnostic workshop will feature leading computer vision researchers and experts from ecology, remote sensing, and the humanities to promote cross-disciplinary discussion of open problems across these fields.

12th Workshop on Medical Computer Vision

Organizers: Alan Yuille, Archana Venkataraman, Can Zhao, Chenyu You, Curtis Langlotz, Dong Yang, Kang Wang, Pedro R. A. S. Bassi, Pengfei Guo, Wenxuan Li, Yang Yang, Yuankai Huo, Yucheng Tang, Yufan He, Yunhe Gao, Yuting He, Zongwei Zhou

Date: 6/3/2026

Time: Full Day

Location: 605



Summary: We propose to host the 12th Workshop on Medical Computer Vision (MCV) at CVPR 2026, building on the strong legacy and momentum of previous years while further expanding its scope and leadership. This year's MCV Workshop will be organized by a broader, multi-institutional team representing Johns Hopkins University, Stanford University, Stony Brook University, NVIDIA, and several other leading institutions. This diverse organizing committee brings together extensive expertise across computer vision, medical imaging, clinical AI, surgical robotics, and embodied AI for healthcare. The workshop will continue its mission of fostering collaboration among academia, healthcare, and industry - engaging students, clinicians, and researchers worldwide. We aim to create an interactive and inclusive environment featuring top-tier invited talks, panel discussions, and networking opportunities that drive forward the frontiers of medical computer vision research and translation.

Sense of Space: Multi-Sensory Modeling for Embodied Intelligence

Organizers: Devin Murphy, Ergys Ristani, Hao-Shu Fang, Jessica Yin, Kaichen Zhou, Kun He, Li Guan, Li Yi, Mengyu Wang, Qi Ye, Rao Fu, Ruoshi Liu, Tomas Hodan, Wojciech Matusik, Xueyan Zou, Yichen Li, Yining Hong, Yiyue Luo, Yunzhu Li, Yuxin Ray Song



Date: 6/3/2026
Time: Full Day
Location: Mile High 2C

Summary: Sense of Space workshop aims to explore multi-sensory modeling to understand our physical world, advancing embodied intelligence in humans and robots. For our first edition, we would like to narrow down our focus to hand-object dexterity manipulation. The workshop addresses key challenges in novel and practical new sensing modalities and sensor design, large-scale, synchronized data acquisition, cross-modal alignment, and generalizable representations for dexterous, real-world embodied tasks. It highlights emerging non-visual sensing modalities, multi-sensory manipulation data acquisition, the transfer of human manipulation skills to robots via large-scale annotated data, and promotes the development of open benchmarks and datasets for reproducible research.

Urban Scene Modeling: Structured, Semantic, and Synthetic 3D Habitats

Organizers: Dmytro Mishkin, Ilke Demir, Jack Langerman, Ruisheng Wang, Shangfeng Huang, Tolga Birdal, Xiang Sean Ma, Yang Wang, Yuzhong Huang



Date: 6/3/2026
Time: Full Day
Location: Mile High 3B

Summary: The 3rd Urban Scene Modeling (USM3D) Workshop at CVPR 2026 focuses on methods that reconstruct, structure, and semantically organize real-world built environments from heterogeneous visual and 3D data. We emphasize scene representations that move beyond raw point clouds and meshes toward high-level, editable, and task-ready models. The workshop includes invited talks, a peer-reviewed paper track, and public challenges on large-scale urban datasets to support benchmarking, reproducible pipelines, and integration across vision, graphics, photogrammetry, and 3D learning. It aims to connect stakeholders across academia, industry, and beyond around scalable, open, and trustworthy urban scene modeling research and deployment practices.

Visual General Intelligence

Organizers: Christian Rupprecht, Hirokatsu Kataoka, Oishi Deb, Shangzhe Wu, Yonglong Tian, Yoshihiro Fukuhara, Ryouyusuke Yamada



Date: 6/3/2026
Time: Full Day
Location: 703

Summary: AGI is said to be an AI capable of replicating human intelligence in every aspect. It's evident intelligence in the visual domain will contribute to the advent, and discussions and preparations toward this goal are cru-

cial within the CVPR community as well. We might ask, should the pursuit of visual intelligence, termed visual general intelligence (VGI), be seen as an extension of current vision research, or does it require a radical leap or paradigm shift? How we should conduct vision research for the next generation, especially directions that don't rely heavily on the language domain, is essential to explore.

4th Workshop on Vision Based Industrial Inspection

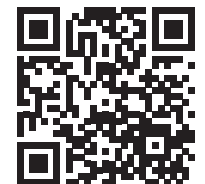
Organizers: Shancong Mou
Date: 6/3/2026
Time: Full Day
Location: 205



Summary: Computer vision (CV) based industrial inspection is a rising field that has gained significant attention from both academia and industry in recent years. However, the synergy between academic research and industry applications remains poor due to the limited exposure of the wider academic community to common issues encountered during real-world production. The VISION workshop series aim to mitigate this gap through invited talks, paper presentations, and a comprehensive industrial inspection dataset release and a data competition. Based on the success of [the 1st VISION workshop](https://vision-based-industrial-inspection.github.io/cvpr-2023/) at CVPR 2023 and [the 2nd VISION workshop](https://vision-based-industrial-inspection.github.io/eccv-24/) at ECCV 2024 and recent [the 3rd VISION workshop](https://vision-workshop.github.io/iccv-2025/) at ICCV 2025, we propose to host the 4th Workshop on Vision-based Industrial Inspection (VISION) at CVPR 2026.

Workshop on Autonomous Driving

Organizers: Alexander Liniger, Chiyu Max Jiang, Dragomir Anguelov, Luc Van Gool, Maying Shen, Nadine Chang, Vincent Casser, John J. Leonard, Jose M. Alvarez



Date: 6/3/2026
Time: Full Day
Location: 603

Summary: The CVPR 2026 Workshop on Autonomous Driving (WAD) brings together leading researchers and engineers from academia and industry to discuss the latest advances in autonomous driving. Now in its 9th year, the workshop has been continuously evolving with this rapidly changing field and now covers all areas of autonomy, including perception, behavior prediction and motion planning. In this full-day workshop, our keynote speakers will provide insights into the ongoing commercialization of autonomous vehicles, as well as progress in related fundamental research areas. Furthermore, we will host a series of technical benchmark challenges to help quantify recent advances in the field, and invite authors of accepted workshop papers to present their work.

The 1st Workshop on Deployment of Foundation Models for Embodied AI

Organizers: Burhaneddin Yaman, Hongyang Li, Selim Engin, Xin Ye, Yingxue Zhou, Yuliang Guo, Yunsheng Ma



Date: 6/3/2026
Time: Full Day
Location: Four Seasons 2

Summary: The 1st Workshop on Deployment of Foundation Models for Embodied AI (WDFM-EA) aims to advance the efficient, scalable, and safe

deployment of large foundation models in embodied intelligence systems – spanning autonomous driving, robotics, and general-purpose embodied agents. Building on the success of our previous Workshops on Distillation of Foundation Models for Autonomous Driving at CVPR & ICCV 2025, which focused on deploying foundation models for self-driving systems, this new edition broadens the scope to holistically address embodied AI from both wheeled (autonomous driving) and legged (robotic) perspectives. The workshop explores synergies between these domains, emphasizing how shared challenges in perception, reasoning, and latency can be jointly addressed through the deployment of multimodal foundation models such as Vision-Language-Action Models and World Models. The 1st WDFM-EA will bring together leading experts from industry and academia to explore deployment practices for scalable adoption of foundation models across all embodied AI applications.

The 2nd 3D-LLM/VLA Workshop: Bridging Language, Vision and Action in 3D Environments

Organizers: David Fouhey, Jiajun Wu, Jianing Yang, Leonidas Guibas, Manling Li, Shengyi Qian, Valts Blukis, Wenbo Hu, Yejin Choi, Yilun Du, Yining Hong, Joyce Chai



Date: 6/3/2026

Time: PM

Location: Mile High 1CD

Summary: The 2nd 3D-LLM/VLA Workshop at CVPR 2026 focuses on the next generation of 3D-grounded Large Language Models (3D-LLMs) and Vision-Language-Action (VLA) systems that unify spatial understanding, reasoning, and interaction. As multimodal LLMs extend beyond text and images to embodied 3D perception, they face fundamental challenges in spatial grounding, memory, temporal reasoning, and action generation. This workshop brings together researchers from vision, language, robotics, and cognition to explore how large models can understand, imagine, and act in the 3D world. Topics include 3D representation learning for LLMs, embodied reasoning, world modeling, instruction following, simulation-to-real transfer, and multimodal evaluation. Through invited talks, paper presentations, and panel discussions, the workshop aims to bridge language-driven understanding and embodied intelligence, establishing a shared foundation for building agents that perceive, reason, and act coherently in complex 3D environments.

10th Affective & Behavior Analysis in-the-wild

Organizers: Alan Cowen, Dimitrios Kollias, Eric Granger, Irene Kotsia, Marco Pedersoli, Panagiotis Tzirakis, Simon L Bacon, Stefanos Zafeiriou



Date: 6/3/2026

Time: PM

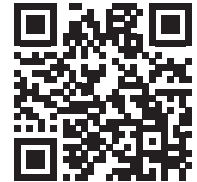
Location: 505

Summary: The 10th Affective & Behavior Analysis in-the-Wild (ABAW) Workshop and Competition at CVPR 2026 is a premier forum for showcasing the latest advances in multimodal analysis, modelling and understanding of human affect and behavior in real-world, unconstrained environments. The workshop highlights cutting-edge methods that integrate facial expressions, head and body movements, gestures, voice, speech and language, supporting both foundational research and impactful applications in human-centered AI. It fosters interdisciplinary exchange across computer vision, machine learning, human-computer interaction, psychology, robotics, healthcare and responsible AI. The programme includes keynote talks from leading experts, technical paper presentations, and the well-established ABAW Competition. This year's competition features

six challenges: valence-arousal estimation, expression recognition, action unit detection, emotional mimicry intensity estimation, ambivalence/hesitancy recognition, and fine-grained violence detection. Built on large-scale in-the-wild datasets, these challenges highlight the importance of multimodal fusion, temporal reasoning, strong pretrained models, and reliable evaluation for understanding human affect and behavior. Overall, ABAW reflects the broader trajectory of the field toward socially meaningful, temporally grounded, and deployment-aware AI systems, while continuing to drive benchmarking, collaboration, and innovation in robust, equitable, and human-centered AI.

AI4RWC: The 2nd International Workshop on Vision Intelligence for Real-world Challenges

Organizers: Alex Zhou, Cedric M John, Daqian Shi, Fausto Giunchiglia, Hao Xu, Honghan Wu, Jishizhan Chen, Xiaolei Diao, Xiongfeng Tang, Yanguo Qin



Date: 6/3/2026

Time: AM

Location: 507

Summary: Bringing visual intelligence into the real world remains a major challenge. Real-world visual data typically exhibits multi-source heterogeneity, scarcity, and imbalance, often with missing values, noise, biases, limited labels, and constraints from complex environments. The AI4RWC workshop takes real-world visual data as its starting point and focuses on building, deploying, monitoring, and continuously improving vision systems in real-world scenarios. Topics include few-/zero-shot and data-centric learning, continual/open-world learning, and knowledge-driven or retrieval-augmented visual reasoning with VLMs/VLLMs. Through keynotes, talks, and panels, it connects vision researchers and domain experts (healthcare, robotics) to foster collaboration toward trustworthy, effective, and impactful real-world computer vision.

The 1st Workshop on AI-assisted Long Video Creation

Organizers: Baohan Xu, Jing Liu, Lisai Zhang, Mingyu Yin, Siqi Wang, Siqian Yang, Xia Ning, Yanzhang Liang, Yidi Wu, Yudong Jiang, Yuxin Hong, Zihao Zhang



Date: 6/3/2026

Time: PM

Location: 712

Summary: This workshop explores how far current AI models are from producing long, multi-shot videos that truly satisfy users and creators. We will bring together researchers, engineers, and creators to discuss advances in long-horizon modeling, controllability, multimodal alignment, and human-AI co-creation. Through talks, papers, and panels, the workshop aims to define benchmarks, evaluation protocols, and design principles for AI-assisted long video creation.

Authenticity & Provenance in the age of Generative AI

Organizers: John Collomosse, Laura Cassani, Luisa Verdoliva, Maneesh Agrawala, Sarah Barrington, Shruti Agarwal, Maty Bohacek

Date: 6/3/2026

Time: PM

Location: 103



Summary: Classical media-forensics tools—often reliant on expert, manual workflows—were designed for smaller scale settings (e.g., adjudicating the authenticity of a single image in court). The rise of generative AI has transformed this landscape: the scale of manipulated and synthetic media has exploded, and new challenges around copyright, provenance, and transparency have emerged. Prior research shows that many existing detectors struggle to generalize and remain robust across models, edits, and real-world distribution shifts. There is no single, silver-bullet solution. The APAI workshop will convene researchers and practitioners from industry who approach mis/disinformation from complementary directions - including multimodal reasoning for media authenticity, fact-checking and attribution methods, community-guided verification, standards-based provenance (e.g., C2PA/CAI), watermarking, and capture-time signals via augmented hardware. Our organizing committee and invited speakers span these threads, enabling collaborations across computer vision, NLP, HCI, security, and policy research.

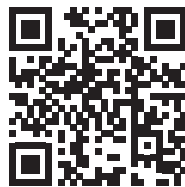
Auto-Annotation with Expert-Crafted Guidelines

Organizers: Jason J Corso, Sara Beery, Shu Kong, Subhansu Maji

Date: 6/3/2026

Time: PM

Location: 711



Summary: Machine-learned visual systems are transforming numerous fields such as autonomous driving, biodiversity assessment, and ecological monitoring, but they hunger for vast, high-quality annotated data. While asking domain experts to manually annotate large-scale data is unrealistic, the current paradigm to scale up data annotation is to have domain experts craft annotation guidelines using visual examples and descriptions for non-expert annotators to apply. This paradigm is commonly adopted by companies which provide data labeling services. Lacking domain knowledge, ordinary annotators often produce annotations that are erroneous, subjective, biased, and inconsistent. Further, this process is labor-intensive, tedious, and costly. This workshop aims to pioneer auto-annotation, developing AI agents that can interpret expert-crafted annotation guidelines and generate labels automatically. In essence, we seek to replace ordinary human annotators with AI.

Cognitive Foundations for Multimodal Models

Organizers: Aditya Chinchure, Gabriel Herbert Sarch, Mohit Bansal, Ranjay Krishna, Sahithya Ravi, Shweta Mahajan, Vered Shwartz, Leonid Sigal, Yossi Gandelsman

Date: 6/3/2026

Time: PM

Location: 610/612



Summary: Despite impressive perceptual and reasoning capabilities, vision-language models (VLMs) face challenges in systematic generalization, sample efficiency, commonsense reasoning, and trustworthy

decision-making. The CogVL workshop provides a forum for researchers across computer vision, natural language processing, and cognitive science to explore how cognitively-inspired frameworks can address these limitations. Our workshop is motivated by the emerging interest in whether cognitive principles such as counterfactual thinking, theory of mind, compositional reasoning, and causal inference can offer a blueprint for more adaptable, robust, and context-aware multimodal intelligence. Our half-day workshop features invited keynote talks, a panel discussion with leading experts, and selected papers. CogVL will also host the BlackSwan Challenge, which evaluates abductive reasoning (inferring hidden causes) and defeasible reasoning (adapting to new visual evidence) in unexpected video events.

Computer Vision for the Built World

Organizers: Iro Armeni, Fuxin Li, Yelda Turkan, Michael Olsen, Marc Pollefeys, Emily Steiner, Sayan Deb Sarkar, Tao Sun

Date: 6/3/2026

Time: PM

Location: 109



Summary: Dynamic, data-rich environments, such as construction sites, offer a unique testbed for advancing spatiotemporal geometric and semantic understanding in computer vision. By capturing how complex 3D scenes evolve over time, we bridge the gap between static “as-designed” models and dynamic “as-built” realities. This workshop explores methods for modeling 3D geometries, dense semantic elements, and temporal dynamics in real-world scenarios. We invite the computer vision community to tackle the challenges of processing large-scale, multimodal data streams, tracking 3D structural changes, and performing continuous semantic reasoning. Through paper submissions, keynote talks, and the Nothing Stands Still construction-data challenge, participants will engage with complex data to advance core vision research in: (i) Spatiotemporal 3D Modeling: Tracking and predicting geometric changes over time; (ii) Multimodal Semantic Understanding: Integrating visual, spatial, and temporal features for dense scene parsing; (iii) Dynamic Scene Reasoning: Bridging static geometric priors with active, evolving real-world states.

Computer Vision with Small Data: Beyond Scale -- Toward Data-Efficient Dynamically-Aware Video Intelligence

Organizers: Chen Chen, James Le, Jason J Corso, Paula Ramos-Giraldo, Shayda Moezzi, Yarong Feng, Octavia Camps, Sarah Ostadabbas, Gedas Bertasius, Baiqi Li, Yuhang He, Zongyi Liu

Date: 6/3/2026

Time: PM

Location: 102/104



Summary: The 3rd CV4Smalls Workshop addresses one of the most persistent challenges in computer vision: building video understanding systems that operate effectively under real-world data scarcity. While large-scale pretraining has revolutionized static image and text understanding, video models remain limited by high annotation costs, complex spatiotemporal dynamics, and poor generalization beyond curated datasets. This workshop calls for a paradigm shift—treating small-data constraints not as barriers but as design principles for robust, physics-aware, and motion-centric video intelligence. By uniting researchers from academia, and industry, CV4Smalls fosters progress in data-efficient representation learning, long-form video reasoning, multimodal fusion, and real-world deploy-

ment. Through technical sessions, invited talks, and a hands-on challenge co-hosted by Voxel51 and Twelve Labs, the workshop aims to establish a roadmap toward generalizable, trustworthy, and dynamics-aware video AI that thrives in unconstrained, real-world environments.

Computer Vision for Biomechanics Workshop

Organizers: Akila Pemasiri, Clinton Fookes, David Ahmedt-Aristizabal, Ethan Goan, Marilyn Keller, Neill D. F. Campbell, David Pagnon, Dario Cazzola, Simon Harrison, Luke Kelly, Glen Lichtwark



Date: 6/3/2026
Time: PM
Location: 112

Summary: Human movement analysis is entering a new era driven by advances in computer vision and machine learning. Accurate estimation of body pose, shape, motion, forces, and joint dynamics is crucial for translating visual data into meaningful biomechanical insights. Yet, challenges such as data variability, limited annotations, and domain generalization persist. This workshop explores emerging intersection of computer vision and biomechanics, bringing together researchers and practitioners to discuss new methods, datasets, and applications that bridge perception and physical understanding enabling precise, data-driven insights and extending the impact of vision-based biomechanics across diverse domains such as rehabilitation, sports performance, and injury prevention.

Sixth Workshop on Neural Architecture Search

Organizers: Amir Atapour-Abarghouei, Elliot J. Crowley, Jingdong Wang, Linchao Zhu, Lukas Neumann, Nik Khadijah Nik Aznan, Ondrej Tybl, Stephen McGough, Teng Xi, Yi Yang, Yifan Sun, David Towers, Wei He



Date: 6/3/2026
Time: PM
Location: 110

Summary: This workshop advances Neural Architecture Search (NAS) for computer vision by bringing together new training-free and data-efficient methods, principled search-space design and fair evaluation under real deployment constraints. The programme blends invited talks, a peer-reviewed paper track, competition and a community discussion on reproducibility, compute cost and ethical use. We target NAS across tasks such as classification, detection and segmentation, spanning edge to foundation-scale settings. A hands-on tutorial lowers the barrier to entry for newcomers. A competition on training-free NAS using a shared search space and blind evaluation rounds out the event.

DataMFM: Emerging Directions in Data for Multimodal Foundation Models

Organizers: Pengyuan Li, Zihan Wang, Zexue He, Xiaolong Luo, Simeng Han, Manling Li, Wenhui Chen, Rogerio Feris



Date: 6/3/2026
Time: PM
Location: 111

Summary: The Emerging Directions in Data for Multimodal Foundation Models (DataMFM) workshop focuses on advancing the data foundations

of multimodal AI. While text data have matured through decades of large-scale curation, visual, audio, and cross-modal data remain limited in quality, diversity, and coverage. This workshop explores emerging directions such as moving from web-scale to world-scale data, agentic and self-generated data pipelines, and high-quality data selection beyond brute-force scaling. Topics include contamination detection, benchmark renewal, and multimodal alignment across modalities. By connecting academia, industry, and open-data efforts, DataMFM aims to define the foundation of next-generation multimodal data ecosystems.

End-to-End 3D Learning

Organizers: Manling Li, Qianqian Wang, Rakesh Ranjan, Siyu Huang, Wenyang Cong, Wuyang Chen, Yan Wang, Zhengzhong Tu, Zhiwen Fan, Dimitris N. Metaxas



Date: 6/3/2026
Time: PM
Location: 501

Summary: End-to-End 3D Learning (E2E3D) focuses on advancing long, brittle 3D pipelines with a single differentiable path from raw sensory streams to geometry, semantics, and action. The workshop covers unified architectures, large-scale pretraining, data curation, and careful evaluation of accuracy, latency, memory, and energy. Case studies span robotics, autonomous driving, XR, and scientific imaging. E2E3D aims to reduce glue code, improve reproducibility, and enable reliable, real-time 3D perception on edge devices. The program features invited talks, posters, and awards, and focuses on reusable design patterns, clear reporting, and open assets to advance 3D research.

3rd Workshop on Efficient and On-Device Generation (EDGE), CVPR 2026

Organizers: Felix Juefei-Xu, Tingbo Hou, Yang Zhao, Haoyu Ma, Enze Xie, Junjiao Tian, Cong Wei, Feng Liang, Zhisheng Xiao, Hui Ding, Xiaoliang Dai, Yanwu Xu, Xuan Ju, Xi Yin, Yuchen Liu, Markos Georgopoulos, Ziqi Huang, Chenfeng Xu, Tao Xu, Lingjuan Lyu, Haolin Jia



Date: 6/3/2026
Time: PM
Location: 203

Summary: Building on top of a successful CVPR 2024 EDGE Workshop and CVPR 2025 EDGE Workshop, the Third Workshop on Efficient and On-Device Generation (EDGE) at CVPR 2026 will continue to focus on the latest advancements in generative AI in the computer vision domain, with an emphasis on efficiencies across multiple aspects. We encourage techniques that enable generative models to be trained more efficiently and/or run on resource-constrained devices, such as mobile phones and edge devices. Through these efforts, we envision a future where these permeating generative AI capabilities become significantly more accessible with virtuous scalability and plateauing carbon footprint.

Second Workshop on Foundation and Generative Models in Biometrics

Organizers: Arun Ross, Hatef Otroschi Shahreza, Luisa Verdoliva, Sébastien Marcel, Vitomir Struc, Zhen Lei

Date: 6/3/2026

Time: PM

Location: 113



Summary: The Second Workshop on Foundation and Generative Models in Biometrics aims to bring together researchers to discuss state-of-the-art advancements, applications, and challenges in using foundation and generative models for biometric recognition, analysis, and security. While foundation models have gained significant attention in recent years, their applications in biometrics remain relatively underexplored. This workshop seeks to encourage discussions that inspire innovation and address the challenges of applying these advanced models in real-world biometric systems. The program will feature invited talks and paper presentations.

1st Workshop on Multi-Agent Robotic Systems: Scaling with Compositional Intelligence

Organizers: Fangchen Liu, Jiayuan Mao, Yilun Du, Yiran Qin, Zhenfei Yin, Nur Muhammad Mahi Shafiullah

Date: 6/3/2026

Time: PM

Location: Mile High 4EF



Summary: MARS workshop aims to bring together researchers across robotics, computer vision, machine learning, and embodied AI to explore how multi-agent collaboration—grounded in physical embodiment, perception, and powered by foundation models—can enable scalable, generalizable, and trustworthy intelligence. It will emphasize vision-based robot-robot coordination, compositional intelligence, and distributed planning across heterogeneous agents. A key focus will be on how visual grounding, spatial reasoning, and multimodal understanding enable agents to perceive shared environments, communicate intent, and cooperate safely. By bridging advances in visual perception, simulation, and real-world deployment, the workshop seeks to establish a unified research agenda for scalable embodied intelligence. Through invited talks, paper sessions, and discussions, MARS aims to foster a focused community advancing the next generation of vision-driven, collaborative embodied systems.

The 5th Workshop on “What is Next in Multimodal Foundation Models?”

Organizers: Eli Schwartz, Hilde Kuehne, Jiaxin Ge, Muhammad Jehanzeb Mirza, Nimrod Shabtay, Roei Herzig, Rogerio Feris, Sivan Doveh, Tsung-Han Wu, Bo Wu, Dantong Niu, David M. Chan, Edson Araujo

Date: 6/3/2026

Time: PM

Location: Four Seasons 4



Summary: Multimodal Foundation Models (MMFMs) have revolutionized AI, achieving remarkable success across vision, language, speech, and beyond. The 5th edition of this workshop aims to explore what is next in this rapidly evolving field, addressing fundamental challenges and charting paths forward. We bring together diverse leaders from academia and industry to discuss critical aspects including model design, training paradigms, generalization, efficiency, ethics, fairness, and open availability.

The workshop features invited talks from distinguished researchers, a call for papers with archival proceedings, poster presentations, and a panel discussion engaging the community on future directions and open problems in multimodal foundation models.

Workshop on Multimodal Human Motion Analysis

Organizers: Arash Ajoudani, Marta Lorenzini, Niaz Ahmad, Olivia Nocentini, Rishabh Dabral

Date: 6/3/2026

Time: PM

Location: 601



Summary: Human Motion Analysis through Multimodal inputs has emerged as an actively researched topic in recent years. This workshop is an attempt to bring the community together and discuss how to unify advances in visual, depth, inertial, and physiological sensing for robust motion understanding across domains and embodiments. There are two focus points of this workshop. The first part focuses on _multimodal perception_, and discusses topics like unified representations, temporal reasoning, and data-efficient learning for action analysis. The second part explores embodied, _human-centered_ intelligence, including topics like foundation models, edge-efficient deployment, as well as responsible evaluation. By bridging multimodal perception and human-centered intelligence, the workshop promotes interdisciplinary dialogue toward robust, generalizable, and ethically aligned frameworks for real-world human motion understanding.

The 1st Workshop on Monitoring the World through an Imperfect Lens

Organizers: Greg Angelides, J. Taylor Perron, Miriam Cha, Nathaniel Maidel, Sara Beery, W. Hamish Mitchell, William T. Freeman

Date: 6/3/2026

Time: PM

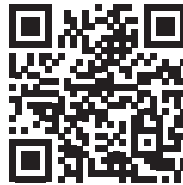
Location: 105



Summary: Multimodal remote sensing has the potential to deliver comprehensive Earth observation by combining complementary sensor capabilities, yet fundamental challenges prevent this potential from being realized. While computer vision has made remarkable progress in multimodal learning with aligned, simultaneously-collected data, remote sensing operates under far more challenging constraints. Satellites collect data asynchronously, at different resolutions, and through fundamentally different imaging physics. For example, synthetic aperture radar (SAR) actively transmits microwave pulses whereas electro-optical (EO) sensors passively capture reflected sunlight. These disparities create a critical gap between theoretical multimodal methods and operational Earth observation systems. This workshop aims to advance multimodal remote sensing beyond simple fusion toward robust methods that embrace real-world constraints.

2nd Workshop on Multimodal Sign Language Recognition

Organizers: Raffaele Mineo, Hamzah Luqman, Simone Palazzo, Maad Alowaifeer, Motaz Alfarraj, Mufti Mahmud, Ahmed Abul Hasanaath, Murtadha Aljubran, Amelia Sorrenti, Federica Proietto Salanitri, Giovanni Bellitto, Concetto Spampinato



Date: 6/3/2026
Time: PM
Location: 106

Summary: MSLR 2026 is the second edition of a rapidly growing venue on multimodal sign language recognition and translation. The program combines invited talks, a peer-reviewed track published in CVPR Workshops, and the SignEval Challenge featuring updated datasets for isolated LIS and continuous SLR. We emphasize privacy-preserving sensing (e.g., radar), healthcare accessibility, and inclusive practices with sign interpreters. Building on the success at ICCV 2025, MSLR 2026 will consolidate a global, interdisciplinary community spanning computer vision, linguistics, healthcare, and Deaf studies.

The 3rd MetaFood Workshop (MTF)

Organizers: Yuhao Chen, Petia Radeva, Jiangpeng He, Bhalaji Nagarajan, Fengqing Zhu



Date: 6/3/2026
Time: PM
Location: 709

Summary: Food presents uniquely complex visual and physical characteristics—it deforms, mixes, and changes through manipulation and consumption. The MetaFood Workshop (MTF) 2026 invites the CVPR community to explore food interaction as a new frontier for embodied perception, video generation, and physics-aware modeling. Understanding how food interacts with tools and humans enables fine-grained video reasoning, not only for estimating how much is eaten but also for revealing intricate multi-material dynamics during cooking and eating. By bridging embodied AI, dynamic 3D reconstruction, vision–language reasoning, and generative modeling, MetaFood 2026 aims to advance physically grounded, fine-grained understanding and synthesis of food in motion.

Machine Unlearning for Vision

Organizers: Alessio Sampieri, Bardh Prenkaj, Bernt Schiele, Fabio Galasso, Iacopo Masi, Joanna Materzynska



Date: 6/3/2026
Time: PM
Location: Mile High 1AB

Summary: Machine unlearning has emerged as a crucial capability for computer vision models that must forget, remove, or steer away from undesired data or concepts. This workshop (in short, MUV) brings together researchers working on unlearning methods for recognition and generative tasks, including selective forgetting, safe image and video generation, privacy-preserving learning, and ethical compliance under regulations such as GDPR and CCPA. By uniting technical advances with responsible AI practices, MUV aims to establish unlearning as a cornerstone of trustworthy visual intelligence. Forgetting is not the opposite of learning but its responsible counterpart.

OpenSUN3D: 6th Workshop on Open-World 3D Scene Understanding with Foundation Models

Organizers: Aikaterini Adam, Alexandros Delitzas, Anna-Maria Halacheva, Ayça Takmaz, Elisabetta Fedele, Federico Tombari, Francis Engelmann, Jan-Nico Zaech, Johanna Wald, Leonidas Guibas, Marc Pollefeys, Nicolai Hermann, Or Litany, Yang Miao, Zuria Bauer, Danda Pani Paudel

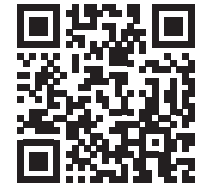


Date: 6/3/2026
Time: PM
Location: 705/707

Summary: For intelligent agents to thrive in the physical world, they must not only perceive but also comprehend and act within it. This year's workshop unites leading researchers advancing world models and spatial intelligence, from 3D perception and scene understanding to generative world representations. Through keynotes and a challenge on open 3D interaction perception, we will explore how machines learn to reason about and engage with their environments. The workshop aims to shape the next generation of spatially grounded AI, bridging scientific discovery, practical deployment, and responsible innovation.

Rediscovering Intelligence: Can AI Still Learn from Humans?

Organizers: Alexey Gavryushin, Asen Nachkov, Chuanyang Jin, Jianwen Xie, Luc Van Gool, Marc Pollefeys, Tianmin Shu, Xi Wang, Yan Zhuang, Yen-Ling Kuo



Date: 6/3/2026
Time: PM
Location: Mile High 4AB

Summary: As AI advances rapidly, it risks losing touch with the cognitive and psychological roots of intelligence. The workshop **“Rediscovering Intelligence: Can AI Still Learn from Humans?”** explores how insights from human cognition, reasoning, and social interaction can inspire more general, grounded, and interpretable AI. We focus on three themes: (1) human-inspired foundations of reasoning and Theory of Mind, (2) learning with humans through feedback and interaction, and (3) going beyond human learning to achieve hybrid intelligence. The workshop bridges AI, cognitive science, and psychology to rethink how machines can still learn from and with us.

Synthetic & Adversarial ForEnsicis

Organizers: Josué Martínez-Martínez, Pooya Khorrani, Danielle Sullivan-Pao, Sheila Alemany Blanco, Giselle Zeno, David Harwath, Bengt J. Borgstrom, Pedro Torres-Carrasquillo, Nicole Meng



Date: 6/3/2026
Time: PM
Location: 107

Summary: Modern machine learning systems must be robust against adversarial data to ensure reliable deployment. Simultaneously, there is a demand to develop authentication systems that can accurately identify synthetic content for the safety of its users. As generative AI and foundation models grow in complexity and become more accessible to the general public, malicious actors exploit model vulnerabilities to create realistic, malicious content, raising security and ethical concerns. This workshop bridges synthetic media forensics with adversarial robustness, focusing on provenance tracking, fingerprinting, and resilient detection methods considering real-time deployed systems requirements. It aims to foster collaboration across the AI, security, and policy domains, develop benchmarks and evaluation tools, and build a shared understanding of joint synthetic-adversarial threats to advance the deployment of trustworthy generative models and the verification of the authenticity of the media.

3rd Workshop on ScanNet++ Novel View Synthesis and 3D Semantic Understanding Challenge

Organizers: Angela Dai, Chandan Yeshwanth, Peter Kotschieder, Yueh-Cheng Liu, Matthias Nießner

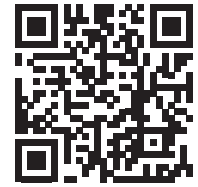


Date: 6/3/2026
Time: PM
Location: 710

Summary: Recent progress in generative modeling and semantic understanding has advanced 3D scene synthesis and interpretation, driving applications in AR/VR, robotics, and computational photography. Yet, large-scale 3D scene understanding and novel view synthesis remain challenging due to the limited fidelity of existing datasets such as ScanNet, Matterport3D, and KITTI, which suffer from low-resolution geometry and visual artifacts. The ScanNet++ benchmark addresses these limitations with high-fidelity 1 mm laser-scan geometry, DSLR imagery, and dense annotations over 1,000 semantic categories, enabling large-scale evaluation of both novel view synthesis and semantic scene understanding. Unlike prior efforts, ScanNet++ evaluates generalization using test views captured outside the training trajectory. This year, we further expand the benchmark and challenge to focus on real-world, commodity-sensor challenges with the introduction of a new iPhone-based Novel View Synthesis benchmark, as well as the release of high-quality 360° RGB-D panoramic data for all benchmark challenges.

Spatial Intelligence for Cultural Heritage

Organizers: Chen Feng, Emanuele Balloni, Fabio Remondino, Jiaxin Lu, Jing Zhang, Livio De Luca, Marina Paolanti, Roberto Pierdicca, Sihang Li



Date: 6/3/2026
Time: PM
Location: 708

Summary: The Spatial Intelligence for Cultural Heritage (SINT4CH) workshop explores the role of AI in preserving and interacting with historical artefacts and sites. By integrating perception, action and learning, the concept of spatial intelligence aims to advance the engagement with cultural heritage (CH). This workshop will explore topics such as 3D reconstruction, AI-driven restoration, and human-AI collaboration, with an emphasis on sustainable and ethical approaches to conservation. With progress in computer vision, robotic interaction, and large-scale language models, participants will explore techniques and future directions that augment human capabilities to ensure the long-term preservation and understanding of our CH. This workshop aims not only to serve as a venue for presenting work in this area but also to build a community and share information in this field.

The 5th Workshop on Transformers for Vision and Multimodal AI

Organizers: Ce Zhang, Di Zhang, Fu-En Yang, Fuxiao Liu, Gedas Bertasius, Han Yi, Jaehun Jung, Le An, Md Mohaiminul Islam, Min-Hung Chen, Mohit Bansal, Rohit Girdhar, Ryo Hachiuma, Seongsu Ha, Shihao Wang, Tianyi Xiong, Yan-Bo Lin, Yulu Pan, Zhiding Yu, Ziyang Wang, Cihang Xie, Guo Chen, Gül Varol, Jose M. Alvarez, Limin Wang



Date: 6/3/2026
Time: PM
Location: 607

Summary: The Transformer architecture has catalyzed a paradigm shift, unifying the fields of computer vision, natural language processing, and beyond. Originally transformative in NLP, its principles now underpin the most powerful foundation models. This includes state-of-the-art models across nearly all vision tasks and a new generation of Multimodal LLMs that seamlessly integrate vision, language, and other modalities. However, this success has brought new challenges. The quadratic complexity of the attention mechanism remains a bottleneck for high-resolution or long-sequence data. Also, the field is actively debating the future of visual backbones: Will Transformers continue to scale effectively? Are emerging alternatives, such as State Space Models (SSMs, e.g., Mamba), more efficient successors? How do we optimally design architectures for unified, multimodal understanding? This workshop will bring together a diverse set of researchers to share insights, debate the limitations of current models, and explore the next generation of architectures for visual recognition and multimodal AI.

The 7th International Workshop and CVML Challenge on Agriculture-Vision: Challenges & Opportunities for Computer Vision in Agriculture

Organizers: Chris Padwick, Ignacio Ciampitti, Jim Ostrowski, Jing Wu, Leandro G. Almeida, Leonardo Lemes Bosche, Mojdeh Saadati, Naira Hovakimyan, Pan Zhao, Paula Ramos-Giraldo, Ripudaman Singh Arora, Humphrey Shi



Date: 6/3/2026
Time: PM
Location: Mile High 2A

Summary: Recent progress in computer vision and deep learning has created exciting opportunities for advancing agriculture. Vision-based technologies can help monitor crops, detect anomalies, and support sustainable practices that improve productivity and resource efficiency. These approaches address challenges of scale, variability, and environmental impact across diverse farming systems. Building on the success of previous editions, the proposed 7th International Workshop and Challenge on Agriculture-Vision at CVPR 2026 will highlight advances in computer vision for agriculture, emphasizing foundation models, multimodal learning, and sustainability. Through research presentations, challenges, and invited talks, the workshop will foster collaboration across academia and related disciplines.

The 2nd Workshop on Test-time Scaling for Computer Vision

Organizers: Yipping Dong, Yichi Zhang, Yu Huang, Shilong Liu, Xueyan Zou, Cihang Xie, Dan Zhang, Miao Liu, Jindong Gu, Lingjuan Lyu, Hang Su, Jun Zhu, Shiguang Shan, Shuicheng Yan



Date: 6/3/2026
Time: PM
Location: 507

Summary: The paradigm of test-time scaling (TTS), which allocates additional computational resources at inference, is rapidly emerging as a critical methodology for enhancing the complex visual reasoning capabilities of vision models. This approach, which allows models to perform deeper, multi-step “thinking” when needed, moves beyond the static capabilities offered by pre-training alone, not only for improving reasoning abilities but also for tackling complex visual tasks even in dynamic environments. However, effectively harnessing this dynamic computation for high-dimensional visual data introduces significant and unique challenges related to computational efficiency, latency, and reliability. This workshop aims to bring together researchers to investigate how the principles of test-time scaling can be applied to computer vision, and to explore novel algorithms and techniques that leverage test-time computation to push the boundaries of computer vision, enabling models to achieve greater accuracy, robustness and interpretability.

The 1st Workshop on Vision for Intelligent Task Assistants

Organizers: Angela Yao, Ehsan Elhamifar, Jason J Corso, Mohsen Moghaddam

Date: 6/3/2026
Time: PM
Location: 108



Summary: Advances in computer vision, multi-modal learning, and AR/VR/XR technologies and smart glasses are converging toward Virtual Intelligent Task Assistants (VITAs)—systems that observe, interpret, and guide humans in complex real-world activities. This workshop bridges computer vision foundations and interactive AR/VR/XR research to enable long-term task understanding and assistance. Topics include learning from long streaming egocentric and exocentric videos, multimodal reasoning, task and step prediction, procedure planning and correction, human-AI collaboration and coaching, and new datasets and benchmarks. By fostering dialogue across disciplines, the workshop aims to define the core challenges and opportunities for building practical and generalizable VITAs.

Thursday, June 4

NOTE: Tutorial rooms are subject to change. Refer to the online site for up-to-date locations. Use the QR code for each tutorial to see its schedule. Here is the QR code for the CVPR 2026 Tutorials page.

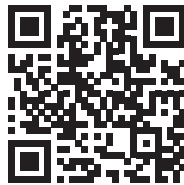


- 7:00 - 17:00 **Registration / Badge Pickup**, Lobby A
- 7:00 - 17:00 **Mother's Room**, Adjacent to Room 102
- 7:00 - 17:00 **Quiet Room**, Adjacent to Room 102
- 7:00 - 17:00 **Prayer Room**, Room 206
- 7:00 - 9:00 **Breakfast**, ExHall C
- 7:30 - 15:00 **Poster Pickup**, ExHall A
- 8:00 - 18:00 **WORKSHOPS / TUTORIALS**
- 8:00 - 18:00 **Posters**, ExHall A
- 10:00 - 11:00 **Coffee Break**, ExHall A
- 12:00 - 13:45 **Lunch**, ExHall C
- 15:00 - 16:00 **Coffee Break**, ExHall A

TUTORIALS

Extending Computer Vision to Hidden Objects: A Tutorial on Millimeter-Wave Imaging and Reconstruction of Occluded Scenes

Organizers: Mingmin Zhao, Laura Dodds
Date: 6/4/2026
Time: AM
Location: 702



Summary: Millimeter-wave (mmWave) sensing is emerging as a new modality for computer vision, enabling perception of objects and scenes that are occluded or invisible to traditional cameras. This tutorial introduces the fundamentals of mmWave imaging, highlighting how its physical properties enable through-occlusion sensing and all-weather perception. It covers both classical signal-processing approaches and recent learning-based methods for 3D reconstruction, segmentation, and scene understanding. The tutorial further provides practical guidance on datasets, tools, and open challenges, offering a comprehensive entry point for researchers interested in extending vision systems beyond visible light.

From Perception to Action: Building Efficient and Deployable Robot Intelligence Pipelines with Open-Source Edge AI Toolkits

Organizers: Samet Akcay, Zhuo Wu, Michael Paulitsch, Ashutosh Kumar, Tao Xiong, Adrian Boguszewski, Sameer Sheorey, Benjamin Ummenhofer, Vladislav Sovrasov



Date: 6/4/2026
Time: PM
Location: 702

Summary: Robotic manipulation has become a key application of embodied AI, but many research pipelines remain difficult to reproduce and deploy in real-world systems. This tutorial presents an end-to-end, open-source workflow for building efficient robot intelligence pipelines, covering data collection, visuomotor policy training, simulation, and deployment on edge hardware. It introduces practical techniques such as teleoperated data acquisition, diffusion- and transformer-based policies, and neural object cloning for simulation-ready assets. The tutorial further emphasizes model optimization and real-time deployment, culminating in a live demonstration of a complete perception-to-action pipeline on an affordable robotic platform.

The Full Stack of Physical AI: Simulation, Foundation Models, and Edge Deployment for Next-Generation Robotics Applications

Organizers: Raymond Lo, Johnny Nunez, Chitoku Yato, Spencer Huang, Mitesh Patel



Date: 6/4/2026
Time: AM
Location: Mile High 2B

Summary: Physical AI systems, including robotics and autonomous platforms, require tightly integrated pipelines spanning data collection, model training, and real-time deployment. This tutorial presents a full-stack perspective on building such systems, covering simulation-based data generation, foundation models for robot control, and deployment on edge hardware. It introduces practical workflows using modern tools for human-in-the-loop data collection, multimodal robot foundation models, and hardware-aware optimization for low-latency inference. The tutorial further highlights challenges in scaling and deploying physical AI systems, providing attendees with actionable guidance and open-source resources for end-to-end robotics development.

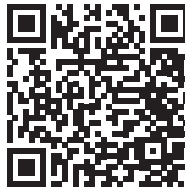
Foundations and Frontiers of Watermarking: Algorithms, Multimodal Extensions, Benchmarks, and Authenticity Frameworks

Organizers: Vishal Asnani, Shruti Agarwal, Benedetta Tondi, Pierre Fernandez, Furong Huang

Date: 6/4/2026

Time: PM

Location: Mile High 2B



Summary: Watermarking has re-emerged as a critical component of trustworthy AI, driven by the rapid growth of generative models and the need for content attribution and authenticity. This tutorial provides a unified overview of watermarking, spanning classical signal-processing foundations and modern deep-learning-based approaches across images, video, audio, and multimodal data. It examines key challenges such as robustness, capacity, and adversarial resilience, along with recent benchmarking efforts and evaluation frameworks. The tutorial further connects these methods to real-world deployment through applications in content provenance, media forensics, and emerging standards such as C2PA, offering a comprehensive perspective on building reliable and transparent media systems.

All You Need To Know About Self-Driving

Organizers: Raquel Urtasun, Abbas Sadat, Sivabalan Manivasagam, Jingkang Wang, Ioan Andrei Barsan

Date: 6/4/2026

Time: Full Day

Location: 301/302



Summary: Autonomous driving has evolved into a complex, full-stack problem that integrates perception, prediction, planning, simulation, and safety within a unified system. This tutorial provides a comprehensive overview of modern self-driving pipelines, covering both traditional modular approaches and emerging end-to-end paradigms. It reviews key components including sensor systems, multi-modal perception, motion forecasting, planning and control, large-scale simulation, and data-centric development. The tutorial further highlights recent advances such as foundation models, generative simulation, and real-world deployment at scale, offering a unified perspective on current challenges and future directions in self-driving systems.

Recent Advances in AI for Medical Imaging: Progress, Challenges, and Future Directions

Organizers: Jiaqi Wang, Peirong Liu, Can Zhao

Date: 6/4/2026

Time: AM

Location: 201



Summary: Artificial intelligence has driven significant advances in medical imaging, improving tasks such as image reconstruction, diagnosis, and clinical decision support across modalities including MRI, CT, X-ray, and pathology. This tutorial provides an up-to-date overview of key paradigms in the field, including physics-informed learning, medical foundation models, and collaborative approaches such as federated and multi-agent systems. It examines major challenges such as generalization, interpretability, data heterogeneity, and privacy constraints, while highlighting emerging solutions and open research directions. The tutorial aims to offer a comprehensive perspective on the development and deployment of reliable, clinically relevant AI systems for medical imaging.

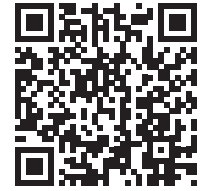
The Road to Convergence: Evolution of Unified Multimodal Models

Organizers: Jindong Wang, Hao Chen, Jiakui Hu, Zhaolong Su

Date: 6/4/2026

Time: PM

Location: 201



Summary: Unified multimodal models are emerging as a new paradigm that integrates understanding and generation across modalities within a single foundation model. This tutorial provides a comprehensive overview of these models, addressing the currently fragmented landscape of architectures, representations, and training strategies. It introduces a unified perspective on key design choices, including modeling paradigms, multimodal tokenization, and alignment methods, while reviewing benchmarks and real-world applications. The tutorial further highlights open challenges such as scalable representation learning and unified world modeling, offering a structured roadmap for future research in multimodal AI.

Computer Vision at Scale: Multi-Camera Tracking, Calibration, and Event Detection for Checkout-Free Retail

<https://checkout-free-tutorial-cvpr26.github.io/>

Organizers: Hareesh Kolluru, Motilal Agarwal, Tanmay Bangalore

Date: 6/4/2026

Time: AM

Location: 203



Summary: Large-scale multi-camera systems are central to real-world computer vision applications, yet their design is shaped as much by infrastructure constraints as by algorithmic advances. This tutorial presents a unified perspective on multi-camera vision through the lens of checkout-free retail, focusing on three core components: automatic camera calibration, real-time multi-object tracking, and structured event detection. It examines how challenges such as asynchrony, partial observability, hardware failures, and edge deployment constraints influence system design and performance. The tutorial further highlights generalizable principles for building reliable, scalable vision systems, bridging the gap between academic methods and real-world deployment.

WORKSHOPS

3D Geometry Generation for Scientific Computing (2nd Edition)

Organizers: Daniel F Martin,
Marissa Ramirez de Chanlatte,
Michael W. Mahoney,
Peter Yichen Chen, Wuyang Chen,
Zhiwen Fan, Chuhan Zou



Date: 6/4/2026
Time: AM
Location: 105

Summary: The fusion of AI and scientific computing is transforming 3D geometry generation for modeling complex natural and engineered systems. Traditional 2D methods fail to capture the multi-scale, heterogeneous data required for accurate simulations in climate science, urban planning, and biomedicine. High-fidelity 3D reconstruction enables deeper understanding and reliable prediction. As scientific challenges grow, from glacier dynamics to resilient infrastructure, scalable 3D methods are essential. Building on the success of 3D4S at WACV 2024, this second edition advances AI-driven 3D modeling to integrate heterogeneous data, optimize for scientific objectives, and balance efficiency with accuracy in real-time reconstruction.

Third Workshop for Learning 3D with Multi-View Supervision

Organizers: Jan Held, Jinjie Mai,
Sara Rojas Martinez,
Silvio Giancola, Abdullah Hamdi,
Gordon Guocheng Qian



Date: 6/4/2026
Time: AM
Location: 703

Summary: The third 3DMV workshop at CVPR 2026 explores the rapidly expanding frontier of 3D and 4D research from multi-view data. Building on the success of previous editions that focused on 3D and object generation, 3DMV 2026 highlights dynamic multi-view datasets, generative 4D models, and their integration with frontier techniques such as LLMs, world models, and video diffusion. A special emphasis is placed on emerging applications in robotics and medical AI, where 3D/4D perception and generation techniques enable robust generalization. 3DMV 2026 aims to unify 2D, 3D, and 4D research communities and foster emerging applications toward scalable, cross-domain multi-view learning.

6th Workshop on 3D Scene Understanding for Vision, Graphics, and Robotics

Organizers: Andreas Geiger, Anpei Chen,
Baoxiong Jia, Chenfanfu Jiang,
Derek Hoiem, Shaofei Wang,
Siyu Tang, Siyuan Huang,
Song-Chun Zhu, Songyou Peng,
Yixin Chen, Yixin Zhu, Chuhan Zou



Date: 6/4/2026
Time: AM
Location: 610/612

Summary: The rapid evolution toward next-generation embodied AI has driven remarkable convergence across computer vision, graphics, and robotics, with growing emphasis on in-the-wild performance and real-world

applications leveraging robust 3D understanding**. Recent breakthroughs have introduced powerful 3D foundational models that demonstrate unprecedented zero-shot generalization across diverse 3D reconstruction tasks. Meanwhile, advances in dynamic reconstruction representations and vision-language-action (VLA) models have enabled practical real-to-sim-to-real transfer, language-grounded robotic manipulation, and navigation in real-world settings. This workshop will focus on exploring how 3D foundational models enable robust, generalizable scene understanding for embodied systems, examining the interplay between geometric reconstruction, semantic grounding, and physical interaction** to advance the next generation of vision, graphics, and robotics research.

Workshop on Any-to-any Multimodal Learning

Organizers: Chenyu Wang, Elisa Ricci, Han Lin,
Roman Bachmann, Sharut Gupta,
Shengqiong Wu, Wei Dai, Yichen Li



Date: 6/4/2026
Time: AM
Location: 502

Summary: The rapid rise of multimodal large models has advanced the integration of language, vision, audio, and beyond. However, most systems remain limited to fixed modality pairs, lacking the flexibility to reason or generalize across arbitrary combinations—unlike humans, who seamlessly integrate diverse sensory inputs. The Any-to-Any Multimodal Learning (A2A-MML) workshop explores the next frontier of multimodal intelligence: building systems that can understand, align, transform, and generate across any modalities. We center discussions on three pillars: (1) Representation learning for generalizable, disentangled multimodal features; (2) Transformation for seamless translation and fusion across heterogeneous modalities; and (3) Collaboration for understanding how modalities cooperate or conflict to achieve coherent reasoning. Bringing together researchers from vision, language, audio, 3D, robotics, and cognitive science, A2A-MML aims to chart a roadmap toward machines that perceive, reason, and communicate across modalities as fluidly as humans.

The 3rd Workshop on New Trends in AI-Generated Media and Security

Organizers: Daniel S. Schiff, Luisa Verdoliva,
Matthew Stamm, Min Chen, Shu Hu,
Siwei Lyu, Xiaoming Liu, Xihui Liu,
Xin Wang, Yung-Hsiang Lu



Date: 6/4/2026
Time: AM
Location: 102/104

Summary: AI-generated media encompasses a wide range of content—including images, videos, audio, and text—produced through advanced artificial intelligence techniques. Recent progress in generative models has brought substantial benefits across diverse domains, enhancing education, healthcare, creativity, business efficiency, and accessibility. Despite these advantages, AI-generated media also introduces serious risks. The same technologies that enable innovation can be exploited to create highly convincing fake news and deepfakes, threatening information integrity, amplifying misinformation, and undermining public trust. The 3rd AIMS Workshop seeks to confront a central question: How can we ensure the responsible, secure, and trustworthy development, detection, and deployment of AI-generated media in ways that align with societal values and sustain public trust?

6th Workshop on CV4Animals: Computer Vision for Animal Behavior Tracking and Modeling

Organizers: Isla Duporge, Mitchell Rogers, Shangzhe Wu, Tuan-Anh Vu, Urs Waldmann, Xiaoxuan Ma

Date: 6/4/2026

Time: AM

Location: 108



Summary: Biological organisms exhibit a wide range of behaviors shaped by both physical and social interactions. Quantifying these behaviors underpins neuroscience, biology, ecology, medicine, behavioral science, and sociology; however, expert-driven ethogramming often lacks objectivity, scalability, and repeatability. Computer vision enables computational measurement, which is especially crucial for in-the-wild biodiversity monitoring, where non-invasive counting and tracking are urgent. However, animal tracking and modeling remain underexplored relative to human-centric vision due to the scarcity of labels, species-specific kinematics (bipedal, quadrupedal, or multipedal), homogeneous textures, and low-resolution remote sensing. This workshop aims to enhance computer vision research by uniting expertise from various fields for sharing, discussion, and collaboration.

2nd Workshop on Computer Vision for Children

Organizers: Daphné Chopard, Ece Ozkan, Ismini Lourentzou, Jintai Chen, Meihuan Huang, Xu Cao, Yifan Shen

Date: 6/4/2026

Time: AM

Location: 106



Summary: While modern computer vision research predominantly centers on adult applications, its transformative potential for children's development, healthcare, and education remains largely untapped. This workshop seeks to bridge this gap by uniting experts from computer vision, pediatrics, psychology, and education to explore the design and impact of advanced computer vision applications tailored for children. Key topics include computer vision algorithms in pediatric healthcare, child and adolescent education, and child psychology, as well as challenges in low-resource settings, crucial ethical considerations, and emerging risks. By fostering this interdisciplinary dialogue, our workshop aims to chart a roadmap towards child-centered computer vision research and broaden the impact of the CVPR community.

The 5th Workshop on Computer Vision in the Wild: Towards Unified Multimodal Agents For Reasoning in the Wild

Organizers: Andrea Tupini, Baolin Peng, Boqing Gong, Hao Cheng, Jianfeng Gao, Lars Liden, Lijuan Wang, Manling Li, Marc Pollefeys, Oier Mees, Qianhui Wu, Qineng Wang, Reuben Tan, Swadheen Shukla, Wenqi Wang, Yu Gu, Yuncong Yang, Zhengyuan Yang, Jianwei Yang, Jiasen Lu, Yong Jae Lee

Date: 6/4/2026

Time: AM

Location: Fours Seasons 2



Summary: We propose to host the 5th Workshop on Computer Vision in the Wild (CVinW): Towards Unified Multimodal Agents for Reasoning in the Wild. This workshop aims to bring together researchers advancing multimodal

AI agents that can perceive, reason, and act in digital and physical environments. Building on recent advances in LMMs and VLAs, we will focus on fostering discussion on crucial capabilities that current multimodal agentic models are limited at. Such capabilities include but are not limited to fine-grained spatiotemporal reasoning, causal inference, long-horizon planning and memory, and robust tool-use. To help measure progress, we also propose to host two challenges: MindCube (spatial mental models under partial observability) and SITE (cross-modal spatial intelligence) with public leaderboards. Finally, along with our invited speakers whose collective expertise spans spatial intelligence, segmentation, robotics, reinforcement learning and multimodal reasoning, the panel discussion and program offer a vision-centric yet cross-disciplinary agenda for multimodal reasoning.

The Second Workshop on the Evaluation of the Generative Foundation Models

Organizers: Karim Bouyarmane, Maria Zontak, Mehmet Saygin Seyfioglu, Suren Kumar, Wisdom Oluchi Ikezogwo, Xu Zhang, Wenyi Wu

Date: 6/4/2026

Time: AM

Location: Mile High 2C



Summary: Generative AI, from content creation with Flux and Veo to real-time interaction with GPT-4o, shows immense potential. However, this power exposes critical flaws like confident hallucinations, systemic bias and lack of realism. As noted by Karpathy, these models optimize for imitation, not correctness. This flaw is compounded as existing static benchmarks become saturated, failing to capture real-world safety or reasoning failures. This evaluation gap is especially concerning as models rapidly expand into complex multimodal domains (video, audio, 3D), a key focus for the CVPR community. Our workshop, EvaGenFM, brings together academia and industry to address this challenge through invited talks, panel discussions, and paper submissions. It provides a vital forum to share best practices, discuss ongoing efforts, and develop new, socially responsible and robust evaluation protocols to ensure the safe and aligned deployment of future foundation models.

Geometry-Free Novel View Synthesis and Controllable Video Models

Organizers: Andrea Tagliasacchi, David B. Lindell, Despoina Paschalidou, Haiwen Feng, Leonidas Guibas, Marcus A Brubaker, Qianqian Wang, Sherwin Bahmani, Siyu Tang, Boyang Deng, Konstantinos G. Derpanis, Lily Goli

Date: 6/4/2026

Time: AM

Location: 607



Summary: This workshop focuses on recent advances in video generative models and their applications in 3D and 4D generation and reconstruction. Topics include camera- and motion-controlled video synthesis, large-scale 3D/4D reconstruction, neural rendering, and generative model-guided pipelines. A central focus is geometry-free novel view synthesis with video diffusion models, enabling spatial control with as little explicit 3D geometry as possible. The program also covers the distillation of temporal models into spatial representations. By highlighting these developments, the workshop aims to chart the path toward more controllable, photorealistic, and efficient generative pipelines that unify video generation with 3D and 4D reconstruction.

Humans of Generative AI

Organizers: Anand Bhattad, David Forsyth, Elissa M. Redmiles, Eunice Yiu, Jaron Mink, Lucy Qin, Sarah Adel Bargal, Shawn Shan, Shiry Ginosar

Date: 6/4/2026

Time: AM

Location: 710



Summary: The design of generative AI and computer vision systems is often guided by technical benchmarks and in-lab evaluations that can differ from their real-world uses. This misalignment, at best, can lead to inefficiencies and, at worst, cause unintended harms in unforeseen contexts. “Humans of Generative AI” recenters attention on the people who use and are affected by these systems. We invite talks and posters from human-centric research that inform the design or evaluation of generative AI and computer vision systems. Through this workshop, we will encourage and develop cross-disciplinary collaboration between computer vision and human-centric researchers, two often disconnected communities.

2nd Workshop on Knowledge-Intensive Multimodal Reasoning

Organizers: Arman Cohan, Manling Li, Xiangliang Zhang, Yilun Zhao, Zeynep Akata, Zhenwen Liang, Wenhu Chen

Date: 6/4/2026

Time: AM

Location: 704/706



Summary: This workshop aims to advance the frontier of multimodal AI systems that can effectively reason across specialized domains requiring extensive domain knowledge. Recent advancements in multimodal AI—combining information from text, images, audio, and structured data—have unlocked impressive capabilities in general-purpose reasoning. However, significant challenges persist when these systems encounter scenarios demanding deep domain expertise in fields such as medicine, engineering, and scientific research. Such contexts require expert-level perception and reasoning grounded in extensive subject knowledge, highlighting the need for specialized strategies to handle domain-specific complexity. Through invited talks, panel discussions, and interactive poster sessions, researchers and practitioners from diverse backgrounds will share the latest developments, ongoing hurdles, and promising future directions. The workshop aims to foster collaboration and stimulate innovation towards the development of next-generation multimodal AI systems in specialized domains.

The 1st Workshop on Low-Level Vision Frontiers with Generative AI, Preference Optimization, and Agentic Systems

Organizers: Bihan Wen, Elisa Ricci, Jian Wang, Kun Yuan, Lanqing Guo, Sina Alemohammad, Weiping Li, Xin Li, Zhibo Chen, Michael S. Brown, Robby T. Tan, Yeying Jin

Date: 6/4/2026

Time: AM

Location: 504



Summary: Low-level vision underpins imaging, graphics, AR/VR, medicine, and robotics, yet current systems face bottlenecks of user experience, optimization, and deployment in the era of Artificial General Intelligence

(AGI). The LoViF workshop targets AGI-oriented low-level vision by catalyzing the convergence of advanced generative AI, preference optimization, and agentic systems. We explore how generative models, LMMs/LLMs, reward modeling, and agent-based reasoning can improve user experience, intelligence, universality, controllability, and interactivity across key tasks, including visual processing, compression, 3D/4D visual representation, perception evaluation, etc. We will host three related challenges and encourage contributions and discussions that propose frontier solutions from an AGI-oriented perspective in the low-level vision.

Multimodal Algorithmic Reasoning Workshop

Organizers: Honglu Zhou, Joshua B. Tenenbaum, Kevin A. Smith, Suhas Lohit, Anoop Cherian, Kuan-Chuan Peng

Date: 6/4/2026

Time: AM

Location: 601



Summary: We plan to gather researchers working in multimodal reasoning, neuro-algorithmic learning, and cognitive models of intelligence to showcase their cutting-edge research, discuss the latest challenges, as well as bring to the forefront problems at the junction of perception and language modeling that are often overlooked but are pivotal in advancing multimodal artificial general intelligence.

Exploring the Next Generation of Data

Organizers: Despoina Paschalidou, Jingde Chen, Maying Shen, Nadine Chang, Jose M. Alvarez

Date: 6/4/2026

Time: AM

Location: 603



Summary: The CVPR 2026 Workshop on Exploring the Next Generation of Data is to gather researchers and engineers across academia and industry to discuss the latest in data-centric methods. Now in its 2nd year, the workshop has been expanded to cover not only standalone data-centric topics (such as data mixture, scaling, selection, mining, generation, causal discovery) but also automatic data loops that can systematically and realistically improve models by combining these data-centric methods. In this full-day workshop, our keynote speakers will provide insights into the ongoing uses of data-centric methods in commercialization, as well as progress in related fundamental research areas. As data-centric approaches tend to be opaque, especially in industry, we will also host an hour long, moderated open discussion to shed light on this topic. Furthermore, we will host a data mining/curation challenge to encourage new advances in the field and will invite authors of accepted workshop papers to present their work.

6th Omnidirectional Computer Vision Workshop

Organizers: Guillaume Caron, Li Guan, Pierre Moulon, Senthil Yogamani, Stefan Milz

Date: 6/4/2026

Time: AM

Location: 711



Summary: Omnidirectional and large field of view (FoV) vision systems—such as fisheye, catadioptric, and 360° cameras—are transforming scene understanding in computer vision and robotics by enabling comprehensive environmental capture with fewer sensors. However, these advances introduce unique challenges, as most existing machine learning and computer vision algorithms are tailored for pinhole cameras. This workshop will ex-

plore the latest challenges and opportunities in omnidirectional vision, fostering discussion on novel methods, datasets, and applications to advance the field and address the limitations of current approaches.

Open-World Vision

Organizers: Abhinav Shrivastava, Andrew Owens, Neehar Peri, Shu Kong, Tian Liu, Yunhan Zhao, Yuxiong Wang

Date: 6/4/2026

Time: AM

Location: 712



Summary: Open-World Vision emphasizes realistic opportunities and challenges in developing and deploying vision systems in the dynamic, vast, and unpredictable real world, which offers abundant data that can both benefit and challenge training and testing. It contrasts the traditional “closed-world” paradigm, which assumes fixed, known data distributions and categorical labels. Models developed in the closed world tend to be brittle when encountering ever-changing and novel scenarios in the open world. Modern algorithms have shifted to an open-world paradigm, such as pretraining foundation models on massive data sourced from the open world (e.g., Internet). While these models show unprecedented performance and strong adaptability, they inherit biases from their open-world pretraining data and still fail in truly novel or underrepresented scenarios. This workshop aims not only to uncover current limitations, potential risks, and emerging opportunities in the open world, but also to solicit solutions that advance the field toward more robust, fair, and adaptable visual systems.

Personalization in Generative AI Workshop

Organizers: Daniel Cohen-Or, Enis Simsar, Ismini Lourentzou, Nupur Kumari, Pinar Yanardag, Tuna Han Salih Meral

Date: 6/4/2026

Time: AM

Location: Mile High 4CD



Summary: Generative AI has revolutionized creativity and problem-solving across domains; enabling the generation of images, text, videos, and 3D content at unprecedented scale. Yet, despite these remarkable advances, personalization remains one of the most challenging and underexplored frontiers. Building systems that understand and adapt to individual users’ preferences, identities, or contexts raises profound technical, ethical, and societal questions; from representation bias and privacy to model alignment and creative ownership. The P13N: Personalization in Generative AI workshop aims to unite researchers, practitioners, and artists from academia and industry to explore these challenges and opportunities. We invite contributions addressing topics such as user modeling, controllability, fairness, interpretability, evaluation of personalization, human-AI co-creation, and responsible deployment of personalized generative models.

PhysHuman: Physically Grounded Human Perception and Modeling

Organizers: Feng Liu, Youngjoong Kwon, Cheng Zhang, Chaitanya Patel, Jun Liu, Thabo Beeler, Xiaoming Liu, Fernando De la Torre

Date: 6/4/2026

Time: AM

Location: 110



Summary: Modern vision and graphics can reconstruct and synthesize humans with high visual fidelity, yet they rarely model how bodies should move under real-world physical constraints. PhysHuman focuses on learning and evaluating physical quantities as first-class targets from video, IMU, and multimodal data, including ground reaction forces, center of mass, joint torques and moments, and contact or friction states. The workshop brings together researchers from computer vision, biomechanics, and simulation, and highlights datasets, evaluation protocols, and applications in sports, clinical assessment, ergonomics, and safe human digital interaction.

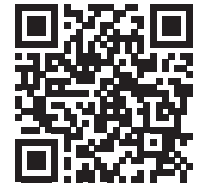
From Perception to Persuasion: Challenges and Advances in Misinformation Detection in Society

Organizers: PRIYANKA SINGH, Pradeep Atrey, Xue Li

Date: 6/4/2026

Time: AM

Location: 113



Summary: Advances in generative vision–language systems have made it easy to fabricate persuasive visual “evidence” and pair it with misleading narratives, accelerating large-scale misinformation campaigns. PP-MisDet focuses on perception and persuasion: how falsified or staged visual content is created, circulated, trusted, and operationalized in high-stakes settings such as elections, health, and conflict. The workshop brings together computer vision, multimedia forensics, human factors, policy, and platform integrity, addressing both algorithmic innovation and interdisciplinary perspectives. We emphasize robustness under coordinated abuse, human–AI collaboration in verification and trustworthy evaluation benchmarks for real-world harm, beyond binary “deepfake vs. real”.

The Eighth Workshop on Precognition: Seeing through the Future

Organizers: Khoa Luu, Nemanja Djuric, Kris Kitani, Utsav Prabhu, Hien Van Nguyen

Date: 6/4/2026

Time: AM

Location: 210/212



Summary: Vision-based detection and recognition studies have recently achieved highly accurate performance, bridging the gap between research and real-world applications. Beyond these research areas, vision-based forecasting will likely be one of the next big research topics in computer vision. This is one of the critical capabilities of humans, and potential success of automatic vision-based forecasting will empower and unlock human-like capabilities in machines. One important application is in autonomous driving, where vision-based understanding of a traffic scene and forecasting of traffic actors is a critical piece of the autonomous puzzle. Another application is the medical domain, allowing deep understanding and prediction of future medical conditions of patients. However, despite its potential and relevance for real-world applications, visual forecasting or precognition has not been the focus of new theoretical studies and practical applications as much as detection and recognition problems. Through the organization of this workshop, we aim to facilitate further discussion and interest within the research community regarding this nascent topic. This workshop will discuss recent approaches and research trends not only in anticipating human behavior from videos but also precognition in other visual applications, such as medical imaging, healthcare, face aging prediction, early event prediction, autonomous driving forecasting, etc.

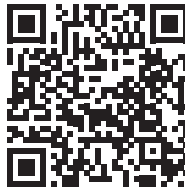
Safe Artificial Intelligence for All Domains

Organizers: Claus Bahlmann, Joachim Sicking, Larissa Triess, MarkusENZweiler, Nikolas Ebert, Oliver Wasenmüller, Peter Schlicht, Stefan Milz, Thomas Stauner

Date: 6/4/2026

Time: AM

Location: 103



Summary: The Safe Artificial Intelligence for All Domains (SAIAD) workshop addresses the development, evaluation, and assurance of safe and trustworthy AI systems across diverse application areas of computer vision and pattern recognition. The workshop covers domains such as robotics, medical imaging, generative AI, and vision foundation models. It provides a platform for researchers and practitioners to present new methods, share best practices, and discuss challenges in designing AI systems that are robust, transparent, and reliable. SAIAD aims to foster cross-domain exchange on methods, tools, and evaluation frameworks along the entire development pipeline from specification to deployment and continuous monitoring.

SPAR-3D: Security, Privacy, and Adversarial Robustness in 3D Generative Vision Models

Organizers: Allen Tu, Ethan Rathbun, Faysal Hossain Shezan, Francis Engelmann, Josué Martínez-Martínez, Nicole Meng, Renjie Wan, Shaoyi Huang, Yingjie Lao, Dongjin Song

Date: 6/4/2026

Time: AM

Location: Mile High 3A

Summary: With rapid advances in computer vision and graphics, 3D scene reconstruction and understanding have become increasingly central to modern visual computing. Methods such as Neural Radiance Fields (NeRF), 3D Gaussian Splatting, 3D diffusion, and vision-language-3D frameworks enable real-time scene rendering and are widely applied in drone operations, cultural heritage, and autonomous driving. However, these systems face security, privacy, and provenance challenges, including adversarial geometry perturbations, data leakage, and model inversion. The SPAR-3D workshop brings together the 3D vision, AI security, and multimodal reasoning communities to advance robustness, traceability, and trustworthy 3D generative systems.



Trustworthy, Robust, Uncertainty-Aware, and Explainable Visual Intelligence and Beyond

Organizers: Arno Blaas, Bhavya Kailkhura, Eunji Kim, Giuseppe Raffa, Kowshik Thopalli, Tsui-Wei Weng, Tammy Riklin Raviv, Trong Nghia Hoang

Date: 6/4/2026

Time: AM

Location: 705/707

Summary: Modern AI systems, particularly those based on vision models and vision-language models, are increasingly deployed in high-stakes domains, yet remain opaque, fragile, and difficult to align across tasks and modalities. Our workshop TRUE-V focuses on trustworthy AI by design: developing transparent, robust, and uncertainty-aware vision and multi-modal systems aligned across the full lifecycle – from training to deployment. The workshop brings together research on interpretability, robustness, uncertainty, and alignment under a principled design paradigm, fostering cross-disciplinary dialogue and community exchange on shared technical and societal challenges, and advancing forward-looking solutions toward reliable, transparent, and safe visual intelligence.



The 8th UG2+ Workshop and Challenge: Bridging the Gap between Computational Photography and Visual Perception

Organizers: Achuta Kadambi, Alex Wong, Dong Lao, Jinwei Ye

Date: 6/4/2026

Time: AM

Location: Mile High 4EF

Summary: UG2+ 2026 bridges computational photography and visual perception to advance robust visual sensing in adverse real-world conditions such as haze, turbulence, low light, and sensor noise. The workshop features invited talks, peer-reviewed papers, and challenges on all-weather image restoration, segmentation under all-weather conditions, and a new task of dynamic object segmentation under atmospheric turbulence and camera motion. By bridging low-level imaging with percep-



tion algorithms, UG2+ promotes dependable vision systems that perform beyond conventional controlled benchmarks. Now in its eighth edition, UG2+ continues to serve as a leading platform for unifying optics, imaging, and computer vision toward readiness for real-world deployment.

Unified Robotic Vision with Cross-Modal Sensing and Alignment

Organizers: Zongwei Wu, Christos Sakaridis, Ilya Chugunov, Yeying Jin, Zhuyun Zhou, Yuedong Tan, Tim Broedermann, Min Liu, Jean Mercat, Xavier Alameda-Pineda, Chao Ma, Tobi Delbruck, Luc Van Gool, Radu Timofte



Date: 6/4/2026
Time: AM
Location: Mile High 4AB

Summary: Modern robotic perception increasingly depends on integrating diverse sensing modalities—RGB, depth, LiDAR, event, audio, and more—to achieve robust understanding in complex environments. While these modalities provide complementary cues, their differences in dimensionality, temporal resolution, and noise make unified perception a persistent challenge. The URVIS workshop brings together researchers from computer vision, robotics, and multimodal learning to explore how unified perception and intelligence can emerge from such diversity. Central questions include: how to extract semantically consistent features from heterogeneous inputs, how to dynamically balance modality reliability, and when fewer sensors can outperform many. URVIS aims to move beyond simple fusion toward adaptive, explainable, and efficient multimodal systems capable of human-level understanding across diverse conditions.

VizWiz Grand Challenge: Interpreting Images and Videos Taken by Blind People

Organizers: Danna Gurari, Jeffrey Bigham, Neelima Prasad, Zhuoheng Li

Date: 6/4/2026
Time: AM
Location: 709



Summary: A common goal in computer vision research is to build machines that can replicate the human vision system; e.g., recognize an object/scene category, locate an object, and read text. A natural grand challenge for the computer vision community is to design such technology to assist people who are blind to overcome their real daily visual challenges. In particular, a desirable next step is for vision applications to empower blind people to learn about their surrounding physical world. Our goal for this workshop is to educate researchers about the technological needs of people with vision impairments while empowering researchers to improve algorithms to meet these needs.

9th International Workshop on Visual Odometry and Computer Vision Applications Based on Location Clues

Organizers: Chandra Kambhampettu, Friedrich Fraundorfer, Guoyu Lu, Nicu Sebe, Yan Yan

Date: 6/4/2026
Time: AM
Location: 506



Summary: Visual odometry (VO) and localization have long been fundamental to computer vision, robotics, and autonomous systems. However, the increasing availability of geo-referenced visual data, from GPS-equipped cameras,

drones, vehicles, and mobile devices, opens new opportunities to integrate location-aware reasoning into visual perception and scene understanding. The workshop aims to bring together researchers to explore how spatial, semantic, and temporal location cues can enhance visual odometry, SLAM, mapping, and broader computer vision tasks. This workshop invites contributions addressing novel frameworks that exploit location clues for improving 3D reconstruction, depth and motion estimation, cross-view localization, and visual place recognition, as well as emerging applications in autonomous driving, AR/VR, and Earth observation. By bridging the gap between pure visual methods and location-informed intelligence, this workshop aims to foster dialogue across vision, robotics, and geospatial communities—advancing the next generation of context-aware visual understanding systems.

4th Workshop on Maritime Computer Vision

Organizers: Alberto Quattrini Li, Arnold Wiliem, Arpita Vats, Benjamin Kiefer, Jan Lukas Augustin, Janez Pers, Jon Muhovič, Josip Šarić, Matej Kristan, Matija Teršek, Mingi Jeong



Date: 6/4/2026
Time: AM
Location: Mile High 1AB

Summary: The 4th Workshop on Maritime Computer Vision (MaCVi) brings together computer vision, robotics, and marine science to advance perception for autonomous surface vessels (USVs) and related domains. MaCVi 2026 features technical papers, keynotes, and five benchmark challenges: vision-chart data association, thermal obstacle detection, LaRS panoptic segmentation, embedded semantic segmentation, and multimodal (RGB-LiDAR-thermal) semantic segmentation. A Generalist Meta Challenge evaluates cross-task generalization. Emphasis is on accuracy and embedded real-time feasibility. By uniting datasets, tasks, and communities, MaCVi accelerates reliable, deployable maritime perception.

The 6th Workshop of Adversarial Machine Learning on Computer Vision: Safety of Vision-Language Agents

Organizers: Aishan Liu, Alan Yuille, Dacheng Tao, Dawn Song, Jiakai Wang, Jin Hu, Jing Shao, Julia Karbing, Juntao Dai, Philip Torr, Tianyuan Zhang, Xianglong Liu, Yinpeng Dong, Zhenfei Yin, Vishal M. Patel, Xinyun Chen



Date: 6/4/2026
Time: Full Day
Location: 708

Summary: Building upon the advances of foundation models, vision-language agents are emerging as a central paradigm for perception, decision-making, and human-AI interaction. However, the growing autonomy and complexity of such agents have amplified their adversarial and safety-critical risks. These vulnerabilities highlight challenges in building safe, robust, and trustworthy vision-language agents for real-world applications in autonomous driving, robotics, medical, and industrial systems. The 6th Workshop on Adversarial Machine Learning in Computer Vision: Safety of Vision-Language Agents aims to bring together researchers and practitioners from computer vision, multimodal learning, and AI safety communities to advance the frontier of trustworthy vision-language agents. Continuing the success of the previous five CVPR AdvML@CV workshops, which have attracted thousands of participants and widespread attention, the 2026 edition will feature keynote talks by leading experts, contributed papers, and an international challenge on adversarial robustness for multimodal agents.

11th Workshop on Computer Vision and Multimodal Microscopy Image Analysis

Organizers: Steven Finkbeiner, Mei Chen, Dimitris N. Metaxas, Oren Kraus
Date: 6/4/2026
Time: Full Day
Location: 112



Summary: As biomedical discovery increasingly relies on the synergy of imaging, multi-omics, and AI, the 11th CVMI Workshop continues its mission to bridge computer vision and the life sciences. Since its CVPR 2016 debut, CVMI has pioneered explainable, quantitative analysis of complex biological systems. This year, we spotlight the transformative power of Large Foundational Models (LFMs) in driving multimodal integration- linking microscopy with omics to tackle high-stakes diagnostic challenges like cancer and pathology screening. Beyond technical innovation, CVMI 2026 emphasizes clinical utility, rigorous evaluation, and interpretability. Featuring world-class invited speakers, highly selective peer-reviewed papers, and work-in-progress spotlights, the workshop remains a leading interdisciplinary forum for AI-driven biomedical breakthroughs.

12th IEEE International Workshop on Computer Vision in Sports

Organizers: Adrian Hilton, Albert Clapés, Anthony Cioppa, Artur Xarles, Graham Thomas, James J. Little, Michele Merler, Rikke Gade, Sergio Escalera, Silvio Giancola, Thomas B. Moeslund
Date: 6/4/2026
Time: Full Day
Location: 503



Summary: Computer vision has recently started to play an important role in sports, in particular for performance optimization and analytics, and in media productions, where computer vision-based graphics in real-time enhances different aspects of the game. The potential of computer vision algorithms in sports is huge, ranging from automatic annotation of broadcast footage, through to better understanding of sport injuries, coaching, and enhanced viewing. The ambition of this workshop is to bring together practitioners and researchers from different disciplines to share ideas and methods on current and future use of computer vision in sports.

The Seventh Annual Embodied Artificial Intelligence Workshop

Organizers: Ade Famoti, Anthony Gerald Francis, Claudia Pérez-D'Arpino, David Hall, German Ros, Mike Roberts, Minyoung Hwang, Oleksandr Maksymets, Peyman Moghadam, Rachith Prakash, Ran Gong, Roberto Martín-Martín, Heewon Kim
Date: 6/4/2026
Time: Full Day
Location: 107



Summary: The Seventh Annual Embodied Artificial Intelligence Workshop (EAI) is a full-day hybrid workshop exploring Embodied World Models, including topics such as Embodied Video and Action Models, Resurgence of Classic Methods, and Long-Horizon Embodied AI. EAI will feature invited speakers, presentations from Embodied AI challenges and their winning

teams, a poster session for accepted papers, topical roundtable discussions, and a concluding debate. This community-building workshop brings researchers together to tackle challenges which force vision-based systems to interact with the real world; see our retrospectives paper <https://arxiv.org/abs/2210.06849> and our website <https://embodied-ai.org/> for more details.

EarthVision: Large Scale Computer Vision for Remote Sensing Imagery

Organizers: Devis Tuia, Hannah Kerner, Jan Dirk Wegner, Loic Landrieu, Nathan Jacobs, Ronny Hänsch
Date: 6/4/2026
Time: Full Day
Location: 507



Summary: The EarthVision workshop focuses on advancing machine learning and computer vision methods for large-scale Earth Observation (EO) data analysis. EO integrates computer vision, machine learning, and signal processing to extract consistent information about the Earth surface from airborne and spaceborne sensors. Key EO tasks include detection, registration, fusion, classification, and regression across multi-sensor, multi-resolution, and multi-temporal data, supporting applications such as mapping, hazard monitoring, and climate analysis. As EO underpins 34 out of 40 UN Sustainable Development Goal indicators, EarthVision plays a vital role in developing automated, scalable solutions for global challenges like deforestation, urbanization, and climate change.

Embodied Reasoning in Action: Workshop and Challenge on Embodied Reasoning for Robotic Manipulation

Organizers: Chan Hee Song, Dieter Fox, Haoquan Fang, Jiafei Duan, Maya Guru, Wentao Yuan, Yejin Kim, Yi Ru Wang, Ziqiao Ma, Jason Ren
Date: 6/4/2026
Time: Full Day
Location: 605



Summary: Embodied reasoning (grounding perception, language, and action in the physical world) is central to robotic manipulation. Foundation models such as LLMs, VLMs, and 3D vision models offer broad knowledge and compositional generalization, yet converting these abilities into reliable closed loop manipulation remains challenging. This workshop unites researchers in robotics, embodied AI, perception, planning, control, simulation, and sim to real to chart the next steps in reasoning centered manipulation. We present two community challenges: RoboSpatial, which tests spatial understanding (where to point, fit, and place) across views and scenes, and PointArena (Point Bench), which measures precise language guided pointing by scoring the selected pixels. Together, these tracks evaluate generalization across viewpoints, sensors, and object sets; robustness to environmental perturbations; and the fidelity of language to action grounding. <https://embodied-reasoning.github.io/>

4th Workshop on Generative Models for Computer Vision

Organizers: Adam Kortylewski, Fangneng Zhan, Jieneng Chen, Christian Theobalt, Alan Yuille

Date: 6/4/2026

Time: Full Day

Location: 205



Summary: The 4th Workshop on Generative Models for Computer Vision (GCV 2026) builds on the success of its CVPR 2023–2025 editions to explore how generative models can advance visual recognition, scene understanding, and simulation. This year's edition highlights emerging directions such as world models—generative systems that learn predictive, structured representations of visual environments. The workshop will feature invited talks from leading researchers, an open-call paper track, and poster presentations. By fostering collaboration between synthesis, recognition, and simulation communities, GCV 2026 aims to shape the next generation of generative methods that both create and understand the visual world.

2nd Workshop on Agents in Interaction, from Humans to Robots

Organizers: Alan Yuille, Dandan Shan, Homanga Bharadhwaj, Jiaman Li, Xiaolong Wang, Yufei Ye

Date: 6/4/2026

Time: Full Day

Location: Mile High 2A



Summary: There has been growing interest in learning robot skills from humans—both of which can be viewed as physical agents that interact with the world. In recent years, computer vision researchers have focused on creating digital twins of humans in virtual environments that behave like real ones, while roboticists have been building physical agents capable of interacting with the real world. We believe that progress in one field can greatly benefit the other. On one hand, virtual humans can be regarded as a special form of robotic agents; on the other hand, robots can learn manipulation and locomotion from human demonstrations, including those performed by simulated humans. Through the proposed workshop, we aim to bring these two fields together to explore common challenges—such as retargeting, embodiment gap, contact modeling, data scarcity, and the role of foundation models, among others. We hope this workshop will serve as a fertile ground for inspiring new research directions and fostering cross-disciplinary collaboration in this space.

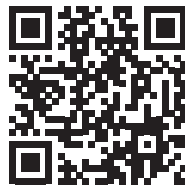
2nd Workshop on Human-Interactive Generation and Editing

Organizers: Alan Yuille, Hengshuang Zhao, Jinbo Xing, Radu Timofte, Scott Cohen, Shaoteng Liu, Tianyu Wang, Xi Chen, Xin Yu, Zhe Lin, Xiaojuan Qi, Yuanhao Cai

Date: 6/4/2026

Time: Full Day

Location: 109



Summary: The rapid evolution of generative AI has reshaped content creation across images, video, and 3D/4D visuals. This workshop focuses on cutting-edge methodologies, practical applications, and open challenges in image/video/3D/4D generation and related editing tasks with an emphasis on flexible and friendly human interactions and multi-modal control signals. And it will also serve as a platform for researchers and practitioners to discuss key topics related to visual content creation and editing with versatile interactions.

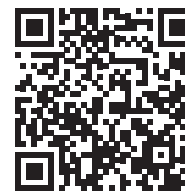
How Do Vision Models Work?

Organizers: Antonio Torralba, David Bau, Rohit Gandikota, Tamar Rott Shaham, Amil Dravid, Yossi Gandelsman

Date: 6/4/2026

Time: Full Day

Location: Mile High 1EF



Summary: How Do Vision Models Work (HOW) is a follow-up to last year's Mechanistic Interpretability for Vision (MIV) workshop, expanding its focus from mechanistic analysis to a broader scientific study of how vision models learn, represent, and reason. As large-scale training has made learned representations increasingly opaque, developing a systematic understanding of these models has become a central scientific challenge. This workshop brings together researchers to investigate the principles and mechanisms underlying the behavior of vision models through observation, experimentation, and theory.

Mobile AI workshop and associated challenges, 6th edition

Organizers: Andrey Ignatov, Radu Timofte

Date: 6/4/2026

Time: Full Day

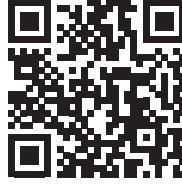
Location: 505



Summary: Over the past years, mobile AI-based applications are becoming more and more ubiquitous. Various deep learning models can now be found on any mobile device starting from smartphones running LLMs, image enhancement, portrait segmentation, face recognition and neural generation models, to IoT platforms performing real-time image classification or smart-TV boards coming with sophisticated image super-resolution algorithms. The performance of mobile NPUs and DSPs is also increasing dramatically, making it possible to run complex deep learning models and to achieve fast runtime in the majority of tasks. While many research papers targeted at efficient deep learning models have been proposed recently, the evaluation of the obtained solutions is usually happening on desktop CPUs and GPUs, making it nearly impossible to estimate the actual performance on real mobile hardware. To address this problem, we propose the 6th edition of the Mobile AI Workshop, where all solutions and deep learning models are evaluated on real mobile neural accelerators such as NPUs, DSPs and TPUs.

Multi-Agent Embodied Intelligent Systems Meet Agentic-AI era: Opportunities, Challenges and Futures

Organizers: Bernadette Bucher, Can Cui, Dongman Lee, Fangzhou Lin, Haibao Yu, Hao Frank Yang, Jiachen Li, Ross Greer, Rui Song, Walter Zimmer, Xiangbo Gao, Yang Zhou, Yue Hu, Yuheng Wu, Yuping Wang, Zhengzhong Tu, Zhiwen Fan, Zilin Huang, Ziran Wang

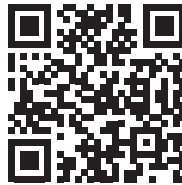


Date: 6/4/2026
Time: Full Day
Location: Four Seasons 1

Summary: This workshop focuses on cooperative intelligence within multi-agent embodied intelligent systems. While individual agents struggle with full autonomy in complex environments, cooperative intelligence enables agents to collaborate with each other for efficient task handling. Recent advances in autonomous driving, robotics, and drone swarms demonstrate the transformative potential of multi-agent cooperation. However, challenges in coordination, data fusion, and real-world dataset acquisition remain. This workshop explores opportunities, challenges, and future directions for multi-agent embodied intelligent systems in the Agentic-AI era through peer-reviewed papers, competitions, and expert talks. Our workshop website is <https://coop-intelligence.github.io>

9th Multimodal Learning and Applications Workshop

Organizers: Benedetta Liberatori, Bodo Rosenhahn, Hao Cheng, Paolo Rota, Pietro Morerio, Vittorio Murino, Michael Ying Yang



Date: 6/4/2026
Time: Full Day
Location: 111

Summary: This workshop addresses the growing importance of multimodal learning, where information from diverse sources, such as images, video, audio, depth, infrared, and sensor data, is integrated to improve perception, reasoning, and generation tasks. We aim to advance research on cross-modal alignment, robust fusion, and adaptation to missing or noisy modalities. Topics include multimodal representation learning, synchronization, dataset design, and evaluation across applications such as autonomous systems, robotics, remote sensing, and human-computer interaction. By fostering collaboration across domains, the workshop seeks to drive progress toward comprehensive, task-oriented multimodal understanding.

11th New Trends in Image Restoration and Enhancement Workshop and Challenges

Organizers: Radu Timofte, Zongwei Wu, Florin Vasluianu, Zhuyun Zhou, Fayaz Ali Dharejo, Marcos Conde, George Ciobotariu, Georgii Perevozchikov, Tim Seizinger, Andrei Dumitriu, Benedikt Hopf, Omar Elezabi, Codruta Ancuti, Cosmin Ancuti, Pierluigi Zama Ramirez, Luigi Di Stefano, Xin Li, Zhibo Chen, Eunpil Park, Yuqian Fu, Xiaohong Liu, Xiongkuo Min, Zheng Chen, Jiatong Li, Kai Liu, Jingkai Wang, Yulun Zhang, Dmitry Vatolin, Longuang Wang, Yingqian Wang, Yulan Guo, Sharif S M A, Qi Jiang, Zhiyuan Li, Bochen Guan, Jie Cai, Guanyi Qin, Jie Liang, Kangning Yang, Lei Zhang, Wenbin Zou, Bin Ren, Yawei Li, Yanan Guan, Lishen Qu, Kun Yuan, Qiufang Ma, Jiebi Yan, Ergor Ershov, Nikola Banic, Lei Sun, Luc Van Gool



Date: 6/4/2026
Time: Full Day
Location: 207

Summary: Image restoration, enhancement, and manipulation are key computer vision tasks aiming at restoring degraded image content, filling missing information, or the needed transformation and/or manipulation to achieve a desired target (with respect to perceptual quality, contents, or performance of apps working on such images). Recent years have witnessed an increased interest from the vision and graphics communities in these fundamental topics of research. Not only has there been a constantly growing flow of related papers, but also substantial progress has been achieved. Each step forward eases the use of images by people or computers for the fulfillment of further tasks, as image manipulation serves as an important frontend. Not surprisingly then, there is an ever growing range of applications in fields such as surveillance, the automotive industry, electronics, remote sensing, or medical image analysis etc. The emergence and ubiquitous use of mobile and wearable devices offer another fertile ground for additional applications and faster methods. This workshop aims to provide an overview of the new trends and advances in those areas. Moreover, it will offer an opportunity for academic and industrial attendees to interact and explore collaborations. 11th NTIRE had 30+ associated challenges gauging state-of-the-art on various tasks.

Video Generative Models: Benchmarks and Evaluation

Organizers: Ashirbad Mishra, Jian Wang, Jinyu Zhao, Kaiyuan Chen, Lin Wang, Mingyang Wu, Pooja Verani, Shuangyu Xie, Shuo Xing, Sicong Jiang, Siyuan Yang, Soumik Dey, Wei Zhou, Yilin Wang, Zhengzhong Tu, Zihan Wang



Date: 6/4/2026
Time: Full Day
Location: Mile High 3B

Summary: Recent progress in video generative models necessitates the development of robust evaluation methods to assess their instruction following, physical realism, human fidelity, and creativity. Existing metrics and current benchmarks are often limited, focusing primarily on semantic alignment while failing to capture crucial imperfections such as structural inconsistencies, unnatural motion, and weak temporal coherence that plague even state-of-the-art generators. Therefore, the VGBE workshop aims to explore **next-generation video evaluators** that are fine-grained, physics plausible, and human-aligned, serving as comprehensive video as-

assessment frameworks capable of reflecting real-world dynamics and identifying subtle imperfections across diverse video generation domains. The workshop further seeks to establish *multi-dimensional*, *explainable*, and *physically grounded* evaluation methodologies that enable **reliable, standardized benchmarks**, so as to facilitating the development of video generative models and supporting their practical deployment in real-world applications.

2nd Workshop on Video Large Language Models

Organizers: Ajmal Saeed Mian, Angela Yao, Ashmal Vayani, C.V. Jawahar, Dmitry Demidov, Gabriele Berton, Garin N. Kessler, Gerard Medioni, Larry S. Davis, Mubarak Shah, Rene Vidal, Rita Cucchiara, Rohit Gupta, Son Tran, Swetha Sirnam, Vimal Bhat, Cees G. M. Snoek, Fahad Shahbaz Khan, Omkar Thawakar



Date: 6/4/2026
Time: Full Day
Location: Four Seasons 4

Summary: This workshop focuses on the latest advancements, challenges, applications, ethical considerations, and future directions of Video LLMs. It consists of keynote talks by leading researchers, a panel discussion on key issues, three challenge tracks that comprehensively assess the diverse capabilities of VidLLMs, and a poster session. Building on this program, the challenge tracks will target compositional video understanding, implicit inference and complex temporal reasoning; the workshop will also emphasize leakage-free datasets, long-video efficiency, and reproducible evaluation, with the goal of enabling practical benchmarks and shared community infrastructure for fair assessment.

Workshop on Visual Concepts

Organizers: Jiajun Wu, Jiayuan Mao, Joy Hsu, Niladri Shekhar Dutt, R. Kenny Jones, Shangzhe Wu, Shenhan Qian, Yunzhi Zhang



Date: 6/4/2026
Time: Full Day
Location: 501

Summary: Visual concept discovery aims to extract compact and structured representations of the visual world. It has played a crucial role in many core problems in computer vision research, including both discriminative and generative tasks. The goal of this workshop is to gather together researchers in computer vision, multi-modal learning, machine learning, and cognitive science to discuss the following topics: - Representations for learning computational models of visual concepts; - Objectives and sources of learning signals to facilitate visual concept learning, incorporating insights from various scientific fields including natural language processing, machine learning, and cognitive science; - Applications of visual concept learning and reasoning, including but not limited to visual scene understanding, robotics, and controllable image generation; - Interpretability of visual learning systems, delving deeper into how these systems learn, represent, and make use of learned concepts in various application domains.

Sight and Sound

Organizers: Andrew Owens, Andrew Zisserman, Antonio Torralba, Hang Zhao, Jean-Charles Bazin, Jiajun Wu, Kristen Grauman, Ruohan Gao, Triantafyllos Afouras, William T. Freeman, Ziyang Chen, Arsha Nagrani

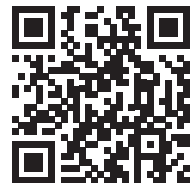


Date: 6/4/2026
Time: Full Day
Location: Mile High 1CD

Summary: In recent years, there have been many advances in learning from visual and auditory data. While traditionally these modalities have been studied in isolation, researchers have increasingly been creating algorithms that learn from both modalities. This has led to a number of exciting developments, such as recent generative models that jointly synthesize synchronized audio and video. The goal of this workshop will be to share recent progress in audio-visual learning, and also to discuss which directions the field should investigate next. To do this, we'll host keynote talks – both from researchers doing audio-visual research, and also from people who study relevant areas of computational audition. We'll also try to summarize recent advances in the field by accepting one-page extended abstracts. Researchers whose abstracts are accepted will have the opportunity to present their work, either in the form of a poster, talk, or multimedia demo.

1st Workshop on Generative 3D Reconstruction

Organizers: Ata Çelen, Daniel Barath, Fabian Manhardt, Federico Tombari, Keisuke Tateno, Marc Pollefeys, Marie-Julie Rakotosaona, Michael Niemeyer, Michael Oechsle



Date: 6/4/2026
Time: PM
Location: 603

Summary: The 1st Workshop on Generative 3D Reconstruction focuses on bridging generative modeling and geometric reconstruction to recover real-world 3D structure from incomplete observations. Unlike purely generative approaches emphasizing visual plausibility, we aim for faithful recovery of the underlying 3D/4D geometry using learned priors, uncertainty reasoning, and physically grounded models. By uniting researchers across neural scene representations, implicit geometry, and 3D Gaussian methods, the workshop defines principles for accurate, trustworthy, and generalizable 3D reconstruction. It will feature invited talks, paper presentations, and discussions shaping the next generation of generative 3D vision methods that reconstruct the real rather than the plausible.

See the World in a Different Light: Physical Appearance Modeling and Relighting in the Age of Generative AI

Organizers: Christian Theobalt, Gordon Wetzstein, Jiajun Wu, Jianchun Chen, Lan Xu, Marc Habermann, Pramod Rao, Valentin Deschaintre, Yingyan Xu, Zhao Dong, xilong zhou, Timo Teufel



Date: 6/4/2026
Time: AM
Location: 210/212

Summary: Recent advances in generative AI, graphics, and vision have led to remarkable progress in relighting, appearance capture and editing, and inverse rendering—spanning scales from small objects to humans and entire scenes. The next major opportunity lies in bridging the gap between the synthesized world and the real world to achieve physically accurate relighting, appearance editing, and capture. Overcoming this challenge will unlock transformative applications in VR/AR, robotics, digital humans, and image, 3D, and video generation and editing. This workshop brings together researchers across these areas to highlight recent advances, discuss open challenges, and explore future directions.

Medical Reasoning with Vision Language Foundation Models

Organizers: Alejandro Lozano, Anas Zafar, Jia Wu, Muhammad Waqas, Nouha Dziri, Tianlong Chen, Xiaoxiao Li



Date: 6/4/2026
Time: PM
Location: 110

Summary: Vision language foundation models demonstrate strong performance in pattern recognition, but lack robust reasoning capabilities, limiting their application in specialized domains. Medical imaging highlights this limitation: diagnosis requires connecting visual findings with clinical knowledge through explicit reasoning. This workshop focuses on advanced reasoning in medical VLMs through interpretable chain of thought, multimodal grounding of visual features in clinical knowledge, evaluation frameworks for reasoning quality, and probabilistic reasoning frameworks for decision making. Through invited talks, contributed papers, panel discussions, and demonstrations, we address reasoning challenges in medical VLMs and establish evaluation frameworks for assessing reasoning quality.

4D Digital Twins: Real-to-Sim-to-Real for Physical AI

Organizers: Amrita Mazumdar, Jonathan Tremblay, Jonghyun Kim, Michael Stengel, Shalini De Mello, Tianye Li



Date: 6/4/2026
Time: PM
Location: Mile High 2C

Summary: Physical AI seeks to imbue embodied agents with the capacity to learn through interaction with their environments, but scaling real-world training data remains a bottleneck. This workshop will explore how high-fidelity 4D neural representations, such as neural radiance fields and Gaussian splatting, can leverage physics-based computer graphics to enable simulation-ready, photorealistic digital twins that support real-to-sim-to-real learning. As part of our workshop, we hold a challenge to generate policies to avoid dynamic, 4D colliders, inspiring the community

to leverage computer vision and graphics techniques towards realistic and physically plausible 4D digital twins for real-to-sim-to-real. By combining graphics' physical grounding with neural models' realism, our workshop aims to bridge the real-to-sim-to-real gap, enabling robust, generalizable foundation models for embodied AI operating in truly dynamic 4D worlds.

2nd Workshop on 4D Vision: Modeling the Dynamic World

Organizers: Andrea Vedaldi, Angjoo Kanazawa, Chen Geng, Gengshan Yang, Jiahui Lei, Leonidas Guibas, Qianqian Wang, Shangzhe Wu, Tarasha Khurana, Lily Goli



Date: 6/4/2026
Time: PM
Location: 506

Summary: We live in a dynamic 3D world. Enabling machines to see and understand the world in 4D (3D + time) unlocks numerous real-world applications across various domains. These include designing autonomous robots capable of navigating and interacting with complex real-world environments, creating immersive and interactive virtual worlds that seamlessly blend with physical reality, and unveiling the mysteries of life and the universe, for instance, in biology and astrophysics. Computer vision lies at the core of these innovations, offering the tools to capture and model the 4D world from partially observed image data. In recent years, we have seen remarkable progress in 3D computer vision, with increasingly robust and efficient models for reconstructing and generating 3D objects and scenes. 4D computer vision (dynamic 3D), as a natural extension of these efforts, is rapidly gaining traction. This workshop aims to continue a dedicated venue for discussions on this topic, bringing together researchers across various domains.

The Third Workshop on Anomaly Detection with Foundation Models

Organizers: Abhishek Aich, Ying Zhao, Kuan-Chuan Peng



Date: 6/4/2026
Time: PM
Location: 712

Summary: The rapid rise of foundation models in healthcare, cybersecurity, industrial inspection, and finance underscores the need to enhance their anomaly detection abilities. Yet, their use in this area remains underexplored. The Anomaly Detection with Foundation Models (ADFM 2026) workshop seeks to bridge this gap by uniting experts to advance research on the intersection of foundation models and anomaly detection. Through discussions of recent breakthroughs and ethical challenges, ADFM 2026 will promote interdisciplinary collaboration and drive the creation of more reliable and effective anomaly detection systems in artificial intelligence.

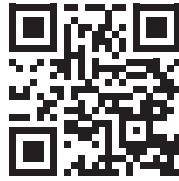
Artificial Intelligence for Space

Organizers: Arunkumar Rathinam, Daniele Gammelli, Djamilia Aouada, Gabriele Meoni, Rajat Talak, Roberto Del Prete, Tat-Jun Chin

Date: 6/4/2026

Time: PM

Location: 710



Summary: The AI4Space workshop at CVPR 2026 will explore the intersection of artificial intelligence and space technologies. As the space sector rapidly expands across Earth observation, astronomy, space robotics, and exploration, AI is emerging as a key enabler for autonomy, onboard intelligence, and large-scale data analysis. The workshop will bring together researchers from computer vision, machine learning, and aerospace to discuss advances and challenges in deploying AI under extreme constraints, bridging domain gaps, and developing reliable, efficient, and trustworthy systems. Through talks and paper presentations, AI4Space aims to accelerate progress toward the next generation of intelligent space missions.

2nd Workshop on GenAI for Storytelling

Organizers: Andrew Shin, Yusuke Mori, Yolo Yunlong Tang, Leo Vitasovic, Haitao Yu, Lin Gu, Xianchao Wu, Hana Kopecka, Kiymet Akdemir, Weijia Wu, Chenliang Xu, Chao Huang

Date: 6/4/2026

Time: PM

Location: 105



Summary: Generative AI has made significant strides in creating visually impressive content, yet it often falls short in producing coherent and engaging narratives. Storytelling, whether in written or visual form, requires consistency and depth across multiple aspects, such as character development, plot progression, and setting dynamics, which current generative models struggle to achieve. This workshop aims to address these limitations by exploring the integration of advanced visual generative models, large language models, and multi-modal AI techniques. Our focus is on advancing the capabilities of generative AI to produce coherent and dynamic narratives that maintain visual consistency and accurately convey story elements, rather than just cool graphics. We seek to push the boundaries of generative AI to create storytelling experiences that are both visually engaging and narratively cohesive, which further expands the applicability of AI to content creation.

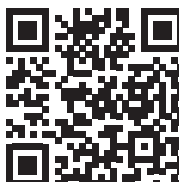
Appearance Understanding and Generation

Organizers: Elena Garces, Giuseppe Vecchio, Iliyan Georgiev, Michael Fischer, Valentin Deschaintre

Date: 6/4/2026

Time: PM

Location: Mile High 4AB



Summary: Understanding and generating appearance are increasingly interdependent challenges. Shape, texture, and reflectance together define how we see the world—and how models should represent it. Advances in intrinsic decomposition, material capture and recognition, and generative rendering now let us analyze and synthesize appearance with unprecedented fidelity. Yet a key question remains: how can a deeper understanding of surface appearance support material perception and interaction, while enabling controllable, identity-preserving editing and consistent multimodal generation? The Appearance Understanding and Generation workshop

unites researchers in vision, graphics, and generative AI to explore representations, datasets, and methods that bridge analysis and synthesis for more interpretable and controllable methods.

Big Model Adaptation In Computer Vision

Organizers: Aida Nematzadeh, Anna Kukleva, Go Irie, Hirokatsu Kataoka, Ivona Najdenkoska, Serge Belongie, Yoshihiro Fukuhara, Ryouosuke Yamada, Yuki M Asano

Date: 6/4/2026

Time: PM

Location: Four Seasons 2



Summary: This workshop explores how large pretrained models are revolutionizing computer vision. We examine emerging techniques where models like Stable Diffusion enable image-to-3D reconstruction, architectures like VGG Transformer (VGGT) leverage pretrained DINO for enhanced visual understanding, and methods like REA harness pretrained models for improved diffusion generation. The workshop covers four key areas: emerging uses of pretrained vision models for 3D and generative AI, prompting techniques for vision models, the role of vision models in multimodal LLMs, and post-pretraining adaptation strategies. BigMAC brings together researchers to address the challenges and opportunities in effectively adapting foundation models for diverse downstream vision tasks.

CVPR 2026 Biometrics Workshop

Organizers: Ajay Kumar, Bir Bhanu

Date: 6/4/2026

Time: PM

Location: 106



Summary: Biometrics accelerates from a niche technology to a foundational infrastructure, powering fintech, national-scale initiatives and welfare programs. Large deployments, such as those in national ID programs and social welfare schemes, underscore its expanding reach and societal impact. At the same time, mainstream adoption has surged: fingerprint and face recognition on everyday smartphones have introduced billions of consumers to seamless, secure authentication. Yet many high-stakes applications still demand accuracy beyond the limits of current techniques. Cutting-edge advances are needed both to push the frontiers within individual modalities and to develop spoof-resistant solutions. This workshop will highlight cutting-edge advances across single- and multimodal systems, featuring premier research from academia, industry, and government laboratories. Alongside keynote talks and expert panel sessions, the program will feature peer-reviewed paper sessions and poster presentations, showcasing innovations that lead to more accurate, resilient, and trustworthy biometric solutions.

Bridging AI and Medical Reality: Computer Vision for Real-world Clinical Translation

Organizers: Jianfei Cai, Jie Liu, Jin Ye, Qiang Ma, Tao Song, Weidong Cai, Wenjia Bai, Xiaoqing Guo, Yicheng Wu, You Zhang, Yunxiang Li, Yutong Xie, Jieneng Chen



Date: 6/4/2026

Time: PM

Location: Mile High 1AB

Summary: The CV4Clinic 2026 workshop aims to bridge advanced computer vision techniques and the complex realities of clinical practice. While CVPR continues to drive progress in imaging, perception, and reasoning, this workshop aims to explore how such advances can translate into real-world medical impacts. By convening experts from academia, industry, and healthcare, CV4Clinic 2026 will focus on key issues of model effectiveness, reliability, and reproducibility in clinical deployment; critically examine the opportunities and challenges of emerging paradigms such as foundation and generative models; and explore future directions for socially meaningful and clinically grounded medical AI.

Computer Vision × Education: Building a Cross-Community Agenda for Multimodal Vision in Classrooms

Organizers: Ekta Sood, Joyce Horn Fonteles, Mariah Bradford, Paul Gavrikov, Prajit Dhar, Janis Pagel, Trisha Mittal, Gautam Biswas, Sidney D'Mello



Date: 6/4/2026

Time: PM

Location: 113

Summary: Computer vision (CV) plays a central role in human-centered AI, yet most models are trained on web-scale benchmarks that poorly reflect real classrooms. Educational data are noisy, private, small-scale, and multimodal (e.g., face, gaze, pose). Students' cognitive/behavioral states (e.g., engagement, mind-wandering) and learning processes (e.g., self-regulation, collaboration) can be inferred from subtle cues in the lab. Still, today's models struggle to generalize to noisy classroom data. CV4Edu brings together computer vision, human-computer interaction, and educational researchers to chart a community agenda for efficient, privacy-aware multimodal data-driven models that work more efficiently and reliably in low-resource, real-world classrooms—potentially launching shared datasets, metrics, and unified practices.

CV4Science: Using Computer Vision for the Sciences

Organizers: David Fouhey, Jing Zhang, Utkarsh Mall, Ye Zhu



Date: 6/4/2026

Time: PM

Location: 709

Summary: AI advancements have become a transformative force, extending beyond their original domain to drive breakthroughs in scientific discovery: an impact highlighted by the 2024 Nobel Prizes in Physics and Chemistry. Computer vision, as one of the core areas in AI research, offers powerful tools for analyzing data, with applications spanning a wide range of scientific fields, from accelerating discoveries in astrophysics to enhancing environmental monitoring and materials science. This workshop aims to bring together researchers working on computer vision and diverse scientific domains to discuss the latest advance-

ments, challenges, and opportunities at their intersections. The goal is to foster interdisciplinary collaboration, build community within the computer vision community, and highlight progress and researchers at the interface of CV and the sciences. CV4Science has been held over the last 2 CVPRs, since 2024. Based on the enthusiastic attendance, we have decided to organize the event again by bringing on speakers and posters from the community to help perpetuate the event.

Domain Generalization: Evolution, Breakthroughs, and Future Horizons (2nd Edition)

Organizers: Biplab Banerjee, Muhammad Haris Khan, Rishabh Lalla, Vineeth N. Balasubramanian



Date: 6/4/2026

Time: PM

Location: 103

Summary: Domain Generalization (DG) is vital for reliable vision under domain, sensor, and temporal shifts. This second edition unifies foundations, systems, and trustworthiness, spanning causal invariances, shift-aware objectives, and reliability metrics (calibration, uncertainty, risk-coverage, OOD) across multimodal, 3D/4D, and embodied settings. Building on the CVPR 2025 first edition, this workshop spotlights agentic/embodied DG, multi-sensor 3D/4D perception, diffusion-driven domain extension and generative test-time augmentation, and privacy-preserving/federated DG. The workshop distills practical recipes, benchmarks, and protocols for deployment. Our invited program features a distinguished, globally diverse roster of speakers across academia and industry and connects theory with real-world impact.

The 2nd CVPR Workshop Proposal on Foundation Models Meet Embodied Agents

Organizers: Canyun Chen, Jiayuan Wu, Jiatao Gu, Jiayuan Mao, Kangrui Wang, Li Fei-Fei, Manling Li, Qineng Wang, Rui Yang, Ruohan Zhang, Weiyu Liu, Wenlong Huang, Yejin Choi, Yining Hong, Yonatan Bisk, Zhiwen Fan



Date: 6/4/2026

Time: PM

Location: 703

Summary: We propose the 2nd Workshop on Foundation Models Meet Embodied Agents (FMEA), advancing how LLM/VLM/VLA-based models enable grounded perception, long-horizon reasoning, and reliable action. We unify vision, robotics, and NLP around an MDP-style pipeline—from goal understanding and subgoal decomposition to action sequencing and transition modeling. We spotlight limiting capabilities: spatial intelligence, physical reasoning, memory/world modeling, planning, and robust tool use across simulation and real settings. We host two challenges: Embodied Agent Interface and EmbodiedBench. The program features invited talks, papers, and a panel, emphasizing ethics, safety, openness, and reproducibility.

The 7th International Workshop on Eye and Gaze in Computer Vision

Organizers: Ales Leonardis, David Wong, Hengfei Wang, Hyung Jin Chang, Michael Stengel, Seonwook Park, Shalini De Mello, Xi Wang, Xucong Zhang, Yihua Cheng, Jean-Marc Odobez



Date: 6/4/2026
Time: PM
Location: 711

Summary: The human eye reflects visual perception, cognitive processes, emotional states, and underlying intentions. Human gaze, as a fundamental manifestation of these processes, serves as a key indicator of attention and intention, holding long-term potential across a broad range of advanced applications, including AR/VR systems, intelligent vehicles, and social behavior analysis. In this workshop, we aim to establish an open platform for the gaze research community to exchange recent advances and explore emerging applications. We welcome researchers from both academia and industry to share their original contributions on eye-related algorithms and applications.

8th International Workshop on Large Scale Holistic Video Understanding

Organizers: Ali Diba, Ehsan Adeli, Joao Carreira, Juergen Gall, Luc Van Gool, Manohar Paluri, Mohsen Fayyaz, Vivek Sharma, David A Ross, Anurag Arnab, Shyamal Buch



Date: 6/4/2026
Time: PM
Location: Mile High 4EF

Summary: The 8th International Workshop on Large-Scale Holistic Video Understanding (HVU 2026) aims to advance research on large-scale, generalist video foundation models that integrate vision, language, and audio understanding. Building on the success of previous HVU editions at CVPR and ICCV, this workshop provides a forum for exploring unified architectures, training methodologies, and evaluation paradigms for video-language models. It brings together researchers and practitioners from

Eighth Workshop on Image Matching: Local Features and Beyond

Organizers: Amy Tabb, Dmytro Mishkin, Eduard Trulls, Fabio Bellavia, Fabio Remondino, Jiri Matas, Luca Morelli, Kwang Moo Yi



Date: 6/4/2026
Time: PM
Location: 504

Summary: Matching two or more images across wide baselines is a core computer vision problem, with applications to stereo, 3D reconstruction, re-localization, SLAM, and retrieval, among many others. Until recently one of the last bastions of traditional handcrafted methods, they too have begun to be replaced with learned alternatives. Interestingly, these new solutions still rely heavily on design intuitions behind handcrafted methods. In short, we are clearly in a transition stage, and our workshop, held every year at CVPR since 2019, aims to address this, bringing together researchers across academia and industry to assess the true state of the field. We aim to establish what works, what doesn't, what's missing, and which research directions are most promising, while focusing on experimental validation.

1st Workshop on Journey to the Awards: Generative AI for Movie-Grade Video Production (J2A), CVPR 2026

Organizers: Felix Juefei-Xu, Stéphane Grabli, Leon Liangyu Chen

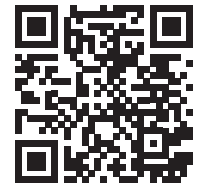


Date: 6/4/2026
Time: PM
Location: Mile High 4CD

Summary: Building on top of the momentum from recent workshops on media generation, we propose the 1st Workshop on Journey to the Awards: Generative AI for Movie-Grade Video Production (J2A) at CVPR 2026. This workshop will explore how generative AI and multimodal LLMs can support the full film production pipeline. Unlike existing media generation workshops, J2A emphasizes studio-caliber cinematic video generation, and the integration of generative AI into the complete movie production pipeline, from ideation to post-production. We aim to bring together researchers and practitioners advancing generative tools that support studios and filmmakers in creating long-form, production-quality work with strong creative coherence. Topics of interest include movie-grade video generation, AI-assisted directing, automatic casting, storyboarding, cinematography, director's cut, editing, post-production, trailer generation, and more. Through these efforts, we envision generative AI advancing the creative industries and opening more pathways to cinematic storytelling.

6th International Workshop on Long-form Video Understanding, Generation and Action

Organizers: Ce Zhang, Gedas Bertasius, Jitendra Malik, Joya Chen, Kristen Grauman, Lorenzo Torresani, Matt Feiszli, Wei Li, Yiqi Lin, Yulu Pan, Zejun MA, Ziyun Zeng, Mike Zheng Shou



Date: 6/4/2026
Time: PM
Location: 704/706

Summary: The LOVEU workshop aims to enhance understanding and foster advancements in long-form video, focusing on building a general multimodal model with rich world knowledge learned from various long-form videos. The workshop will explore the unique challenges and emerging methodologies for enabling video models to understand, generate, and act upon long-form video data. Special focus will be placed on scaling video models to reason, generate, and act over long temporal contexts to address complex real-world tasks.

The 2nd Workshop on Multi-Modal Reasoning for Agentic Intelligence

Organizers: Amita Kamath, Anda-Raluca Epure, Annie S Chen, Mahtab Bigverdi, Philip Torr, Ranjay Krishna, Yijiang Li, Zhenfei Yin, Zixian Ma, Alexander T Toshev



Date: 6/4/2026

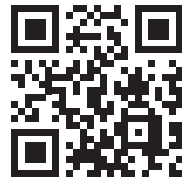
Time: PM

Location: Mile High 3A

Summary: AI agents built on Large Language Models (LLMs) already demonstrate solid reasoning on tasks such as coding, analysis, and knowledge-intensive inquiry. With the emergence of Multimodal Foundation Models (MFMs), these agents can now fuse visual, language, and audio signals, enabling far richer perception and more informed decisions. This workshop brings together researchers working on the next generation of multimodal reasoning for AI agents—spanning vision–language–audio modeling, embodied and virtual agents, OS/screen agents, AI co-scientists, and evaluation. We will discuss architectures for native multimodal reasoning, methods for semantic alignment and uncertainty modeling, and strategies to reduce the computational overhead of image / video-heavy inputs. We will also highlight applications in robotics, HCI, and automated discovery, where agents must integrate perception with long-horizon decision-making. The goal is to chart a research agenda for robust, efficient, and trustworthy multimodal agents that operate in complex, open-ended environments.

Pixel-level Video Understanding in the Wild Challenge

Organizers: Elisa Ricci, Jiayu Miao, Nikhila Ravi, Philip Torr, Si Liu, Song Bai, Yi Yang, Yi Zhu, Yunchao Wei, Zongxin Yang, Cees G. M. Snoek, Henghui Ding



Date: 6/4/2026

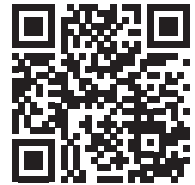
Time: PM

Location: 502

Summary: The 5th PVUW workshop focuses on advancing pixel-level video understanding in the wild, with an emphasis on dense prediction tasks such as semantic, instance, and panoptic video parsing, dense tracking and video object segmentation, as well as multi-modal or omnimodal video understanding. It particularly encourages research addressing real-world challenges in complex scenes, such as frequent object disappearance-reappearance, severe occlusions and crowding, low-light and adverse weather conditions, multi-shot transitions, camouflaged objects, and knowledge-dependent reasoning. This year's edition introduces upgraded datasets and challenge tracks in complex scenes to promote the development of robust and deployable solutions for real-world pixel-level video understanding.

4D World Models: Bridging Generation and Reconstruction

Organizers: Aashish Rai, Angela Dai, Angela Xing, Chaerin Min, Jonathon Luiten, Lingjie Liu, Nikolaos Sarafianos, Rao Fu, Xiaowei Zhou, Xiaoyan Cong, Yiqing Liang, Zekun Li, Aayush Prakash, Fernando De la Torre, Srinath Sridhar



Date: 6/4/2026

Time: PM

Location: 203

Summary: The realistic reconstruction and generation of dynamic virtual worlds is an essential task for scene understanding and robot learning. Recent advances in NeRF and 3DGS have driven the progress of reconstruction and generative models in 3D/4D. However, 3D/4D generation and reconstruction have largely evolved as separate research tracks, despite their conceptual overlap and potential for synergy. Integrating these paradigms can offer a wider perspective to identify optimal representation and scalable data strategies for building 4D world models. This workshop aims to bring together researchers to explore the unified frameworks of understanding, reconstructing, and generating the 3D/4D world.

Third Workshop on Simulation for Autonomous Driving

Organizers: Yiyi Liao, Maximilian Igl, Ekaterina Tolstaya, Azadeh Dinparastdjadid, Jingkang Wang, Kashyap Chitta, Maximilian Naumann, Shuhan Tan, Zan Gojcic, Sanja Fidler, Shimon Whiteson



Date: 6/4/2026

Time: PM

Location: 102/104

Summary: On-road testing of autonomous vehicles is costly, making simulation essential for developing autonomous driving. Beyond evaluation, simulation also plays an increasing role in closed-loop training and reinforcement learning. This workshop brings together researchers from multiple disciplines to advance both fronts of the simulation challenge: improving fidelity and leveraging simulation for training and evaluation. Topics include sensor simulation, behavior modeling, reinforcement learning, and closed-loop evaluation. The program features keynotes, contributed papers, and a panel discussion. The workshop will also host a challenge on closed-loop evaluation using photorealistic simulators to better benchmark the real-world performance of end-to-end driving systems.

Second Workshop on Skilled Activity Understanding, Assessment & Feedback Generation

Organizers: Angela Yao, Basura Fernando, Paritosh Parmar, Brendan Tran Morris



Date: 6/4/2026

Time: PM

Location: 705/707

Summary: Imagine a world where computer vision-based systems can analyze a video of an athlete, a surgeon, a patient, or a factory worker and instantly provide expert-level actionable feedback-correcting techniques, identifying inefficiencies, and helping people refine their skills in real time. Thanks to rapid progress in video understanding, this vision is becoming reality. AI-powered systems can now analyze complex human activities, assess performance, and generate intelligent feedback, unlocking new possibilities in sports, healthcare, manufactur-

ing, education, rehabilitation, and beyond. Through Expert Keynotes and Invited Contributions, this workshop will explore the cutting edge of skilled activity understanding, assessment, and feedback generation, bridging research and real-world applications. As AI grows more capable of modeling human expertise, it unlocks new possibilities for personalized coaching, democratized skill development, and scalable training. This workshop brings together researchers and practitioners to advance how AI perceives and enhances human skills and ability.

ScaleBot: The First Workshop on Scalable Robot Learning Systems

Organizers: Hengshuang Zhao, Philip Torr, Shenyan Gao, Sijin Chen, Xihui Liu, Yang Gao, Yingdong Hu, Yuke Zhu, Yuxiang Lu, Zhenfei Yin, Jianfei Yang

Date: 6/4/2026

Time: PM

Location: 610/612



Summary: The First Workshop on Scalable Robot Learning Systems (ScaleBot) aims to bring together researchers and engineers from computer vision, natural language processing, robotics, and related fields to discuss how to build scalable learning systems for general-purpose robots. Inspired by the success of large-scale foundation models in vision and language, this workshop focuses on scaling the entire robot learning pipeline – covering topics including data collection, training recipes, model architectures, system integration, and evaluation. By fostering interdisciplinary dialogue and collaboration, ScaleBot seeks to identify key challenges, share recent advances, and outline future directions toward scalable, robust, and generalizable robot systems that can effectively operate in the real world.

The 3rd Workshop on Synthetic Data for Computer Vision

Organizers: Jieyu Zhang, Ranjay Krishna, Rundong Luo, Shobhita Sundaram, Wei-Chiu Ma, Weikai Huang, Zixian Ma

Date: 6/4/2026

Time: PM

Location: 607



Summary: The workshop aims to explore the use of synthetic data in training and evaluating computer vision models, as well as other related domains. During the last decade, advances in computer vision have been catalyzed by the release of meticulously curated human-labeled datasets. Recently, people have increasingly resorted to synthetic data as an alternative to labor-intensive human-labeled datasets for its scalability, customizability, and cost-effectiveness. Synthetic data offers the potential to generate large volumes of diverse and high-quality vision data, tailored to specific scenarios and edge cases that are hard to capture in real-world data. However, challenges such as the domain gap between synthetic and real-world data, potential biases in synthetic generation, and the generalizability of models trained on synthetic data remain. We hope that the workshop can provide a forum for discussion and encouragement of further exploration in these areas.

Visual Anomaly and Novelty Detection - 4th Edition

Organizers: Guansong Pang, Jawad Tayyub, Latha Pemula, Paul Bergmann, Paula Ramos-Giraldo, Philipp Seeböck, Samet Akcay, Lars Heckler-Kram, Toby P. Breckon

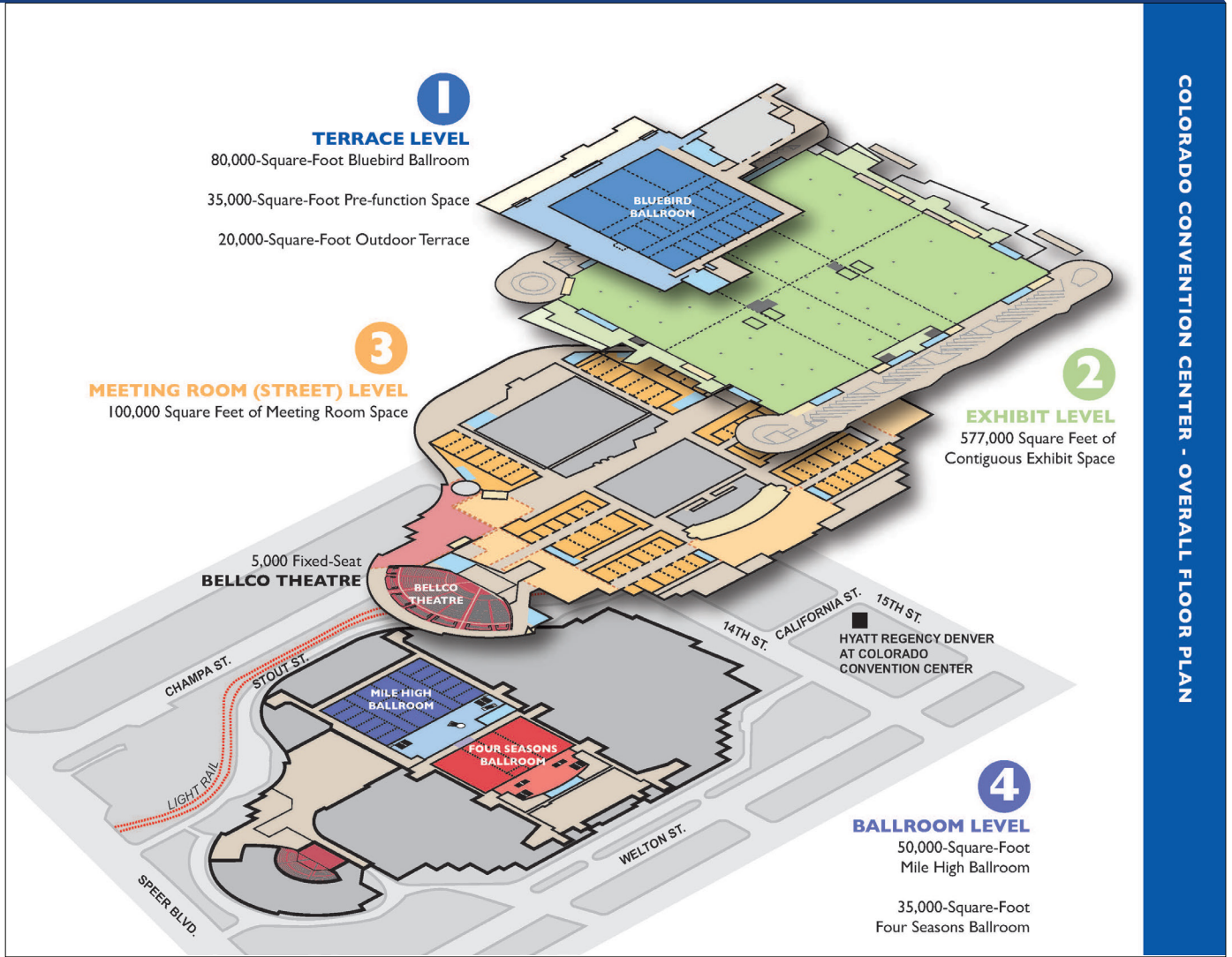
Date: 6/4/2026

Time: PM

Location: 601



Summary: Anomaly detection—also known as novelty or out-of-distribution detection—identifies samples that fall outside the known distribution of visual data. Detecting what “doesn’t belong” is critical for robust real-world systems in industrial inspection, medical imaging and beyond, yet remains unsolved due to the open-ended nature of anomalies and their absence during training. Building on three prior editions, the Visual Anomaly and Novelty Detection (VAND) workshop at CVPR 2026 unites research across supervised, semi-, and unsupervised paradigms, including few-, one-, and zero-shot learning, and features challenges on robust and efficient segmentation and VLM-based logical and structural anomaly detection.



BALLROOM LEVEL

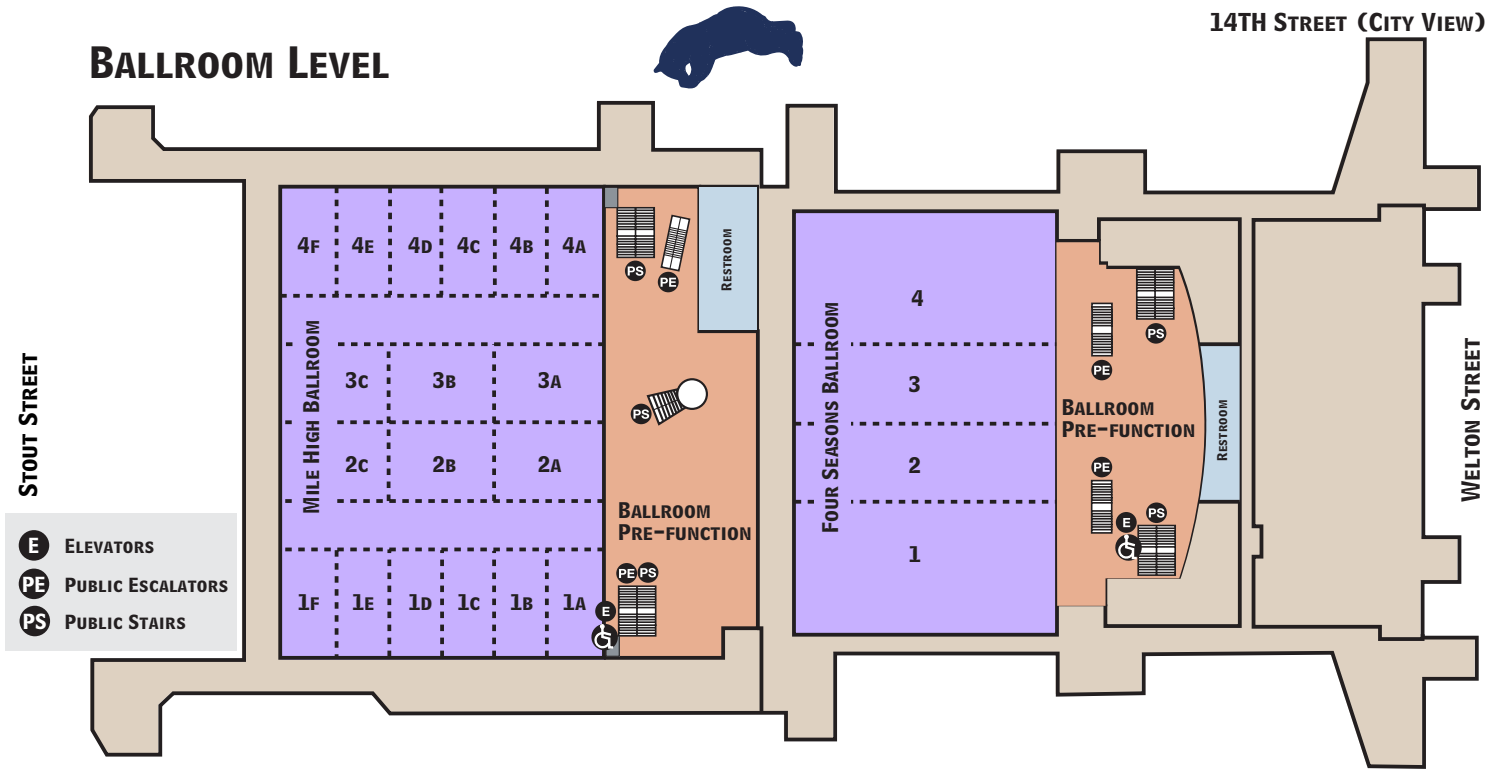


EXHIBIT HALLS

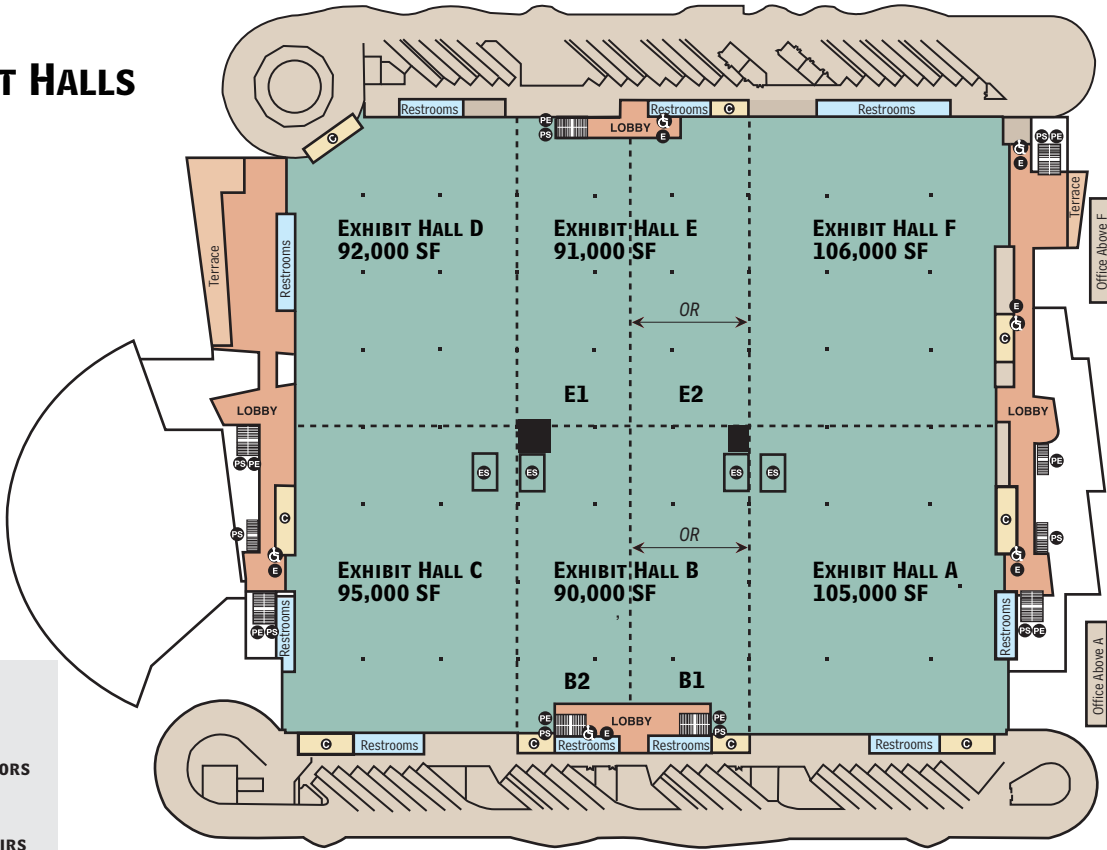
SPEER BLVD (MOUNTAIN VIEW)

14TH STREET (CITY VIEW)

CHAMPA STREET

WELTON STREET

- C** CONCESSIONS
- E** ELEVATORS
- PE** PUBLIC ESCALATORS
- PS** PUBLIC STAIRS
- ES** EMERGENCY STAIRS



MEETING ROOM

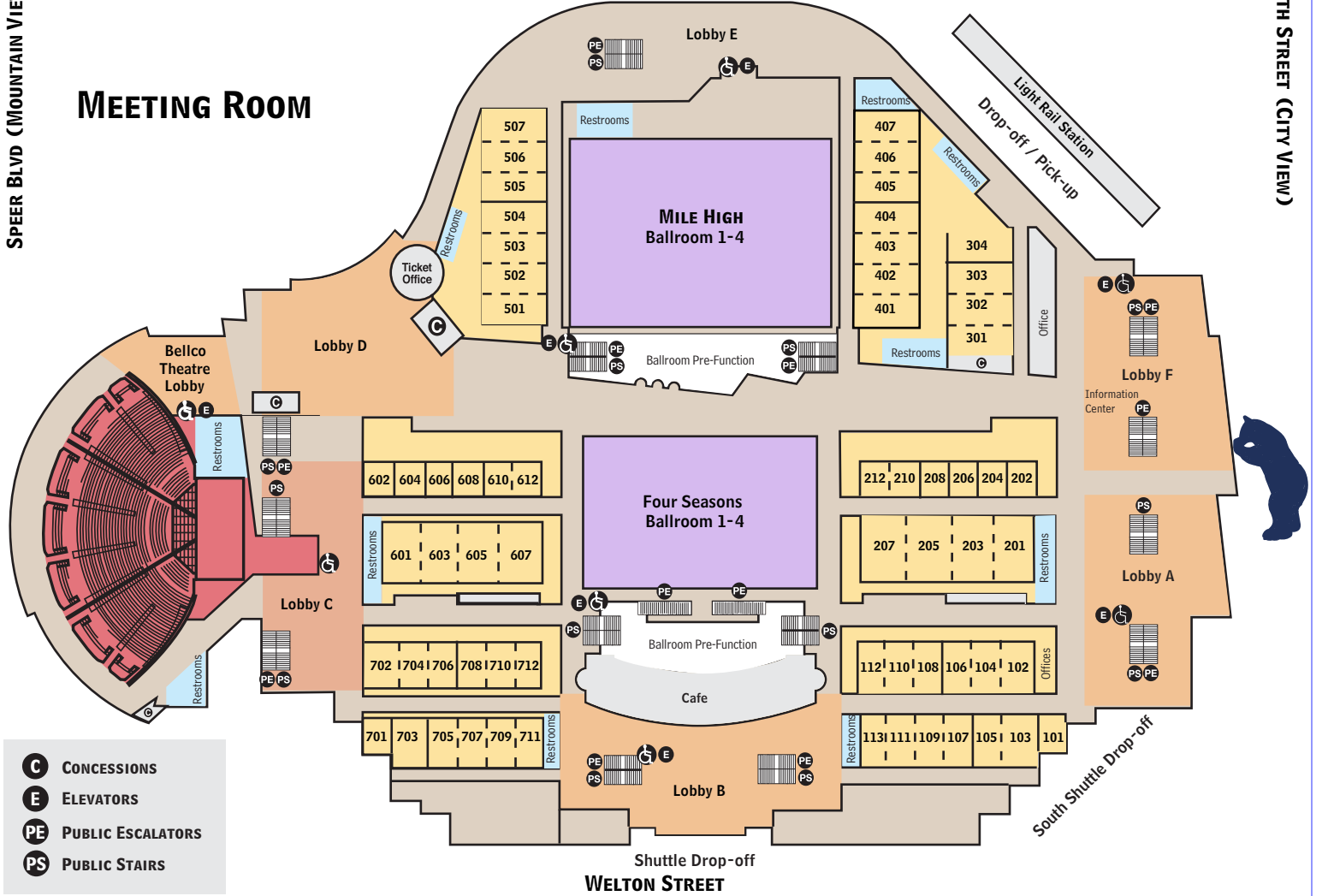
SPEER BLVD (MOUNTAIN VIEW)

14TH STREET (CITY VIEW)

STOUT STREET

WELTON STREET

- C** CONCESSIONS
- E** ELEVATORS
- PE** PUBLIC ESCALATORS
- PS** PUBLIC STAIRS





CVPR
JUNE 3-7, 2026



DENVER
COLORADO